Int. J. Sci. R. Tech., 2026 3(11)
A Multidisciplinary peer-reviewed Journal
www.ijsrtjournal.com [ISSN: 2394-7063]

Original Article

A Review of Machine Learning-Based Dyslexia Detection Using fMRI Data

Adarsh Pradhan[1*], Shikhar Kumar Sarma[2], Mirzanur Rahman[2]

[1] *Department of Computer Science and Engineering, Girijananda Chowdhury University, Guwahati, Assam, India.*
[2] *Department of Information Technology, Gauhati University, Guwahati, Assam, India.*

ABSTRACT

An individual with good intelligence and proper education may still suffer from dyslexia, making reading and spelling an arduous task for them. Although behavioural assessments are commonly used to diagnose dyslexia, they fail to reflect the brain's neurological complexity. Functional Magnetic Resonance Imaging (fMRI) offers a non-invasive means to observe brain activity, revealing characteristic patterns in regions associated with language and phonological processing in individuals with dyslexia. When it comes to early detection of dyslexia, fMRI and machine learning (ML) integration can play an important role. In this review, we explore the application of various machine learning techniques, including Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), and Convolutional Neural Networks (CNNs) to fMRI data in the context of dyslexia research. We highlight key studies that have leveraged these methods to identify distinctive brain activity patterns, achieving promising classification performance. Despite advancements, challenges persist, such as data scarcity, high dimensionality, model interpretability, and computational demands. Moreover, the use of predictive models in human populations also raises ethical issues. Future directions, such as the creation of Explainable Artificial Intelligence (XAI), the use of few-shot learning and transfer learning for small datasets, and the incorporation of multimodal data to increase detection accuracy, are also discussed. This interdisciplinary field, where neuroscience, computer science, and medical imaging intersect, holds transformative potential for the early detection and diagnosis of dyslexia.

Keywords: Dyslexia, Functional Magnetic Resonance Imaging (fMRI), machine learning (ML), Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs)

INTRODUCTION

It is estimated that 10% of people worldwide suffer from dyslexia and struggle with reading and writing [1]. Despite normal cognitive ability and educational opportunity, individuals with dyslexia often struggle with word recognition, spelling, and phonological decoding [2]. These difficulties are persistent and can significantly impact academic achievement, self-esteem, and long-term socioeconomic outcomes. Traditionally, dyslexia is diagnosed through behavioral assessments, which may be prone to subjectivity. Recent advances in neuroscience have offered critical insights into the neural basis of dyslexia. Functional Magnetic Resonance Imaging (fMRI) detects brain activity by measuring changes in blood oxygenation (BOLD signals) [3]. Research has shown that individuals with dyslexia exhibit under-activation in left-hemispheric regions involved in language and phonological processing, such as the inferior frontal gyrus, temporoparietal cortex, and occipitotemporal area [4,5,6]. Such results support the idea that neuroimaging could provide markers for identifying dyslexia.

However, the analysis of fMRI data poses significant challenges. Each scan yields tens of thousands of voxel time series, creating a high-dimensional, noisy dataset with complex temporal dynamics. Extracting meaningful patterns from such data demands advanced computational tools—this is where machine learning (ML) plays a transformative role. ML algorithms can detect non-linear, multivariate relationships in large datasets and learn discriminative patterns associated with clinical conditions like dyslexia [7,8]. The integration of ML with fMRI in dyslexia research is particularly promising. It enables automated feature extraction, pattern recognition, and

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

classification of neuroimaging data. Common tasks include identifying brain regions that distinguish dyslexic individuals from typical readers, classifying individuals based on their neural signatures, and understanding functional connectivity differences. Models such as Support Vector Machines (SVMs) [9], Convolutional Neural Networks (CNNs) [10,11], and hybrid transformer-based approaches [12,13] have been successfully applied to these tasks. Furthermore, medical image analysis, especially in neuroimaging, has seen rapid adoption of ML for various diagnostic applications, from Alzheimer's disease to ADHD [8,14]. The same principles are now applied to reading disorders. Notably, fMRI provides the ability to capture both task-related and resting-state brain activity, offering complementary perspectives on neural dysfunctions in dyslexia.

Despite this progress, challenges remain. These include the scarcity of large, high-quality datasets specific to dyslexia, the "black-box" nature of deep learning models, generalization difficulties due to overfitting, and ethical concerns regarding data use in pediatric populations. These barriers highlight the need for interdisciplinary efforts to develop explainable, generalizable, and ethically sound ML models for neuroimaging-based dyslexia detection.

This review aims to:

- Summarize recent research that applies ML to fMRI for dyslexia classification.
- Identify technical, methodological, and ethical challenges and explore future directions, including explainable AI, multimodal fusion, and few-shot learning.

2 Existing Research for Dyslexia Detection

2.1 Traditional ML – SVMs, ANNs, PCA, and ICA

Support Vector Machines (SVMs). To effectively process the complex, non-linear, and high-dimensional data in fMRI images, Support Vector Machines (SVMs) can employ kernel functions and handle such complexities [9]. SVMs can readily identify the optimal boundary between groups, like dyslexic and non-dyslexic, and by applying kernels, SVMs can handle non-linear decision boundaries, making them well-suited for the complex patterns found in brain scans. In the study by Tamboer et al. [15], SVMs and k-Nearest Neighbors (k-NN) were used on structural MRI data. They extracted grey and white matter volumes and used these as inputs. The SVM was then trained to differentiate between dyslexic and typical readers, achieving up to 82% accuracy. Similarly, Cui et al. [16] employed SVMs on white matter features from diffusion tensor imaging (DTI) and reported strong classification performance (accuracy: 83.61%), linking disrupted connectivity with reading difficulties.

Artificial Neural Networks (ANNs). Artificial Neural Network (ANN) is another model that learn non-linear mappings between input and output spaces [17]; and with the right regularization methods like dropout [18], can generalize well. García Chimeno et al. [19] combined functional Magnetic Resonance Imaging (fMRI) and Diffusion Tensor Imaging (DTI) data with traditional ML. A variety of classifiers (LDA, SVM, ANN, k-means clustering, etc.) were trained on these data (with/without PCA) to distinguish between dyslexic and control subjects. The best model, a three-layer ANN, without PCA, achieved 94.87% accuracy, 94.73% sensitivity, and 95% specificity.

Principal Component Analysis (PCA) and Independent Component Analysis (ICA). Given the high dimensionality of fMRI, Principal Component Analysis (PCA) [20] and Independent Component Analysis (ICA) [21] are used as feature selection or dimensionality reduction techniques and often paired with SVM or ANN classifiers. For dimensionality reduction and decorrelation, PCA converts input data into a series of components ordered by their explained variance. Conversely, ICA separates mixed data, such as brain activation sources activated during different cognitive tasks, into statistically independent constituent parts. McKeown et al. [21] used ICA to identify task-related brain activity components, which were later input into classifiers for group comparison. PCA, as discussed in Jolliffe and Cadima [20], is often the first step in reducing data to a more manageable size before applying learning algorithms.

2.2 Deep Learning Methods

Convolutional Neural Networks (CNNs). CNNs offer automated feature learning from raw neuroimaging data, bypassing the need for manual feature selection. CNNs are particularly effective in capturing local spatial patterns in volumetric data like fMRI [10,11]. In the work by Zahia et al. [11], three parallel 3D CNNs were constructed with batch normalization and dense layers. These models were trained on task-based fMRI scans obtained during linguistic tasks, enabling

them to learn spatial activation patterns linked to dyslexic brain function. The system achieved a classification accuracy of 72.73%, specificity of 71.43%, precision of 60%, and F1-score of 67%. Silva et al. [22] also utilized CNNs applied to task-based fMRI data but added Grad-CAM interpretability. 2D convolutional neural networks via grammar-based genetic programming were used, getting the best CNN accuracy of 95%, far outperforming a linear SVM (70%). They also attempted to showcase the parts of the brain that are impacted by dyslexia using Grad-CAM.

2.3 Multimodal and Transformer-Based Methods

Multimodal Fusion (anatomical and functional MRI). By combining data from multiple sources, models can develop a richer understanding of the disorder. For instance, Harismithaa and Sadasivam [12] developed a system that fused anatomical and functional MRI using CNN-LSTM hybrids. The time-distributed CNN captured spatial features, while the LSTM modeled temporal sequences, allowing for dynamic brain-state modeling that outperformed single-modality systems.

Functional Connectivity Matrices and Graph Representations. Taran et al. [23] used functional connectivity matrices, derived from pairwise correlations across 264 brain regions, as input features to their classifier models, achieving the noteworthy accuracy of 79%, sensitivity of 92%, and specificity of 64%. Their research revealed that there are neurobiological variations in the brains of dyslexics.

Vision Transformers with Bayesian Optimization on multimodal data—structural MRI, handwriting images, and EEG. Sait and Alkhurayyif [13] integrated features from multimodal data—structural MRI, handwriting images, and EEG, and used specialized vision transformers with a multi-attention fusion layer. They also optimized the XGBoost classifier via Bayesian hyperparameter tuning, producing a notable classification accuracy of 99.8%.

Table 1 shows a tabular representation of some of the research that has been conducted on Machine learning-based dyslexia detection.

Table 1: Existing Research for Dyslexia Detection.

Study	Method	Technique/Tool	Outcome
Tamboer et al. [15]	Traditional ML	SVM, k-NN on structural MRI	Acc: 73–82%
Cui et al. [16]	Traditional ML	SVM on DTI connectivity	Acc: 83.61%, Sens: 75.00%, Spec: 90.91%, AUC: 0.86
García Chimeno et al. [19]	Traditional ML	LDA, SVM, ANN, k-means, etc	Acc: 94.87%, Sens: 94.73%, Spec: 95%,
Jolliffe [20], McKeown [21]	Feature Reduction	PCA and ICA	Improved classifier performance
Zahia et al. [11]	Deep Learning	3D CNN on fMRI volumes	Acc: 72.73%, Spec: 71.43%, Prec: 60%, F1: 67%
Silva et al. [22]	Deep Learning + Interpretability	2D CNN + Grad-CAM	Acc: 85.71% + interpretability via visual heatmaps
Harismithaa & Sadasivam [12]	Multimodal Deep Learning	CNN and Time-Distributed CNN-LSTM (MRI fusion)	Multimodal model outperformed unimodal
Taran et al. [23]	Connectivity-based Detection	Classifier on functional connectivity	Acc: 79% Sens: 92% Spec: 64%
Sait & Alkhurayyif (2025) [13]	Transformer Architecture	Vision Transformer + XGBoost (Structural MRI + handwriting images + EEG)	Acc: 99.8%

3 Challenges and Future Directions: Applying Machine Learning for Dyslexia Detection

3.1 Data Scarcity and Generalization

One of the most pressing limitations in applying ML to dyslexia research is the scarcity of large, well-annotated fMRI datasets. Collecting high-quality neuroimaging data, especially from children with dyslexia, is time-consuming, expensive, and ethically sensitive. As a result, many studies are constrained to relatively small sample sizes, increasing the risk of overfitting and reducing the generalizability of models [11,15,22]. Moreover, discrepancies in imaging protocols, scanner hardware, and participant demographics across datasets make it difficult to reproduce results or transfer trained models between institutions. Although OpenNeuro[24] and other platforms like the Autism Brain Imaging Data Exchange (ABIDE) [25] and the WU-Minn Human Connectome Project [26] provide neuroimaging data, there is a significant gap in this research domain due to the lack of dyslexia-specific fMRI scans.

Transfer learning and few-shot learning could be a viable strategy to handle this issue. By leveraging models pre-trained on large, general neuroimaging datasets, these techniques can adapt knowledge to smaller dyslexia-focused datasets with improved efficiency and accuracy [14,16]. Meta-learning approaches further enhance generalization by enabling models to learn how to learn from minimal examples.

3.2 Need for Interpretability in "Black Box" Models

Deep learning models, particularly CNNs and transformers, have been successful in decoding complex neural patterns. However, they are often thought of as opaque "black boxes," offering little insight into why the different sections of the complex deep learning algorithm made a decision, especially to professionals from a background other than Computer Science. This lack of transparency is a significant barrier to clinical trust and acceptance, especially in neurodevelopmental disorders like dyslexia.

Emerging tools like Grad-CAM [27] have helped visualize important brain regions contributing to a model's output. The field of explainable AI (XAI) can be strongly explored to create models that are not only accurate but also interpretable in neuroscientific and clinical terms [22,23]. Future models should provide visualisations and confidence intervals that correspond with known patterns of brain activity, which should make them easier for medical practitioners and other professionals to use.

3.3 Computational Complexity and Standardization Gaps

Training deep learning models on fMRI data is computationally intensive, often requiring powerful GPUs and large-scale memory resources. This poses a challenge for smaller research teams and institutions without access to such infrastructure [11,13]. In addition, the preprocessing of fMRI data—motion correction, normalization, artifact removal—must be standardized to ensure reproducibility. Tools like fMRIPrep [28] and statistical frameworks described by Friston et al. [29] have helped bring consistency to preprocessing, but more widespread adoption of such tools is needed. Uniform pipelines not only improve reproducibility but also enable easier collaboration and data sharing across sites.

3.4 Ethical Responsibility in AI-Driven Diagnosis

The processing of patient information, such as neuroimaging data, by algorithms raises ethical concerns, such as safeguarding patient information, ensuring informed consent, and preventing diagnostic errors. Furthermore, there is an inherent issue with machine learning models during training: they may contain biases pertaining to demographic inequalities, socioeconomic status, or language background [30]. To ensure ethical integrity, models should be rigorously audited for bias and fairness. We need to ensure ethical design principles are integrated into AI pipelines during their initial development stages.

3.5 Prospective Approach: Multimodal Integration

To improve performance and overcome the limitations of unimodal imaging, researchers are increasingly adopting multimodal integration strategies. Combining fMRI with EEG, structural MRI, handwriting scans, or even genetic data allows for richer and more reliable feature extraction [12,13,31]. Techniques such as graph-based modeling and cross-modality transformers are particularly promising in this area, enabling the model to learn relationships across different data types and scales. Just as importantly, the field needs community-wide standards for

dataset creation, annotation, benchmarking, and reporting. Shared evaluation protocols, open-source model repositories, and collaborative benchmark challenges would greatly accelerate progress while ensuring transparency and accountability.

The challenges as well as the potential solutions for the advancement of dyslexia detection through neuroimaging and machine learning are shown in Table 2.

Table 2: Challenges and Future Directions for Dyslexia Detection.

Challenges	Future Directions/ Solutions
Limited dyslexia-specific neuroimaging datasets	Use of public repositories (e.g., OpenNeuro); creation of specialized datasets; apply transfer learning and few-shot learning from general neuroimaging corpora
Poor model generalization across sites	Development of domain-adaptive models; encouraging cross-site validations; use of meta-learning to improve generalizability
Lack of model interpretability	Integration of explainable AI tools like Grad-CAM; prioritize transparent decision-making frameworks aligned with neural anatomy
High computational complexity	Optimizing model architectures for efficiency; promoting the use of cloud-based or distributed computing platforms
Non-standardized preprocessing and pipelines	Encouraging the adoption of community-endorsed tools like fMRIPrep; enforce reproducibility through open protocols
Ethical issues: privacy, bias, and consent	Conducting bias audits; implementing clear consent and data governance protocols; ensuring interpretability for non-technical stakeholders
Unimodal data limitations	Promotion of multimodal integration: combining fMRI, EEG, DTI, handwriting scans, genetic data for comprehensive insight
Data inefficiency in training	Adopting few-shot learning, meta-learning, and pre-training on large neuroimaging datasets before fine-tuning on dyslexia data

CONCLUSION

When it comes to early dyslexia detection, we require an effective scientific approach that outperforms conventional diagnostic techniques and assists those who suffer from dyslexia. In this study, we emphasize that machine learning in conjunction with early detection capabilities in fields such as neuroimaging, namely functional MRI (fMRI), could provide an alternative solution. We see that several machine learning models, from the traditional Support Vector Machines (SVMs) to the more advanced 3D Convolutional Neural Networks (CNNs) and transformer-based systems, have been used to classify dyslexic brain activity with notable accuracy. Deep learning models, in particular, show strong potential for detecting nuanced brain patterns, while explainable AI (XAI) methodology could help make these predictions more transparent and meaningful.

Nevertheless, significant challenges remain. Data scarcity, generalization issues, model transparency, and ethical concerns, particularly when working with children, are critical barriers that must be addressed. Collaborations across neuroscience, computer science, and clinical fields will be critical to address these gaps. Advances in neuroscience and AI hold the power to reshape dyslexia detection and intervention, offering earlier decisions, individualized support, and more accurate detection for those affected.

Disclosure of Interests. The authors of this paper state no conflicts of interest.

REFERENCES

1. Shaywitz, S.E.: Dyslexia. The New England Journal of Medicine, 338 (5), 307–312 (1998).
2. Lyon, G.R., Shaywitz, S.E., and Shaywitz, B.A.: A definition of dyslexia. Annals of Dyslexia, 53 (1), 1–14 (2003).
3. Ogawa, S., Lee, T.M., Kay, A.R., and Tank, D.W.: Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proceedings of the National Academy of Sciences, 87 (24), 9868–9872 (1990).
4. Pugh, K.R., Mencl, W.E., Jenner, A.R., Katz, L., Frost, S.J., Lee, J.R., and Shaywitz, B.A.: Functional neuroimaging studies of reading and reading disability (developmental dyslexia). Mental Retardation and Developmental Disabilities Research Reviews, 6 (3), 207–213 (2000).
5. Richlan, F., Kronbichler, M., and Wimmer, H.: Meta-analyzing brain dysfunctions in dyslexic

children and adults. NeuroImage, 56 (3), 1735–1742 (2011).

6. Norton, E.S., Beach, S.D., and Gabrieli, J.D.: Neurobiology of dyslexia. Current Opinion in Neurobiology, 30, 73–78 (2015).
7. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A., Ciompi, F., Ghafoorian, M., Laak, J.A.W.M.V.D., van Ginneken, B., and Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88 (2017).
8. Shen, D., Wu, G., and Suk, H.I.: Deep Learning in Medical Image Analysis. Annual Review of Biomedical Engineering, 19, 221–248 (2017).
9. Cortes, C. and Vapnik, V.: Support-vector networks. Machine Learning, 20 (3), 273–297 (1995).
10. Lecun, Y., Bengio, Y., and Hinton, G.: Deep learning. Nature, 521 (7553), 436–444 (2015).
11. Zahia, S., Garcia-Zapirain, B., Saralegui, I., and Fernandez-Ruanova, B.: Dyslexia Detection Using 3D Convolutional Neural Networks And Functional Magnetic Resonance Imaging. Computer Methods and Programs in Biomedicine, 197, 105726 (2020).
12. Harismithaa, L.R. and Sadasivam, G.S.: Multimodal screening for dyslexia using anatomical and functional MRI data. Journal of Computational Methods in Sciences and Engineering, 22 (4), 1105–1116 (2022).
13. Sait, A.R.W. and Alkhurayyif, Y.: Lightweight hybrid transformers-based dyslexia detection using cross-modality data. Scientific Reports, 15, 17054 (2025).
14. Suk, H.I., Lee, S.W., and Shen, D.: Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. NeuroImage, 101, 569–582 (2014).
15. Tamboer, P., Vorst, H.C.M., Ghebreab, S., and Scholte, H.S.: Machine learning and dyslexia: Classification of individual structural neuroimaging scans of students with and without dyslexia. NeuroImage: Clinical, 11, 508–514 (2016).
16. Cui, Z., Xia, Z., Su, M., Shu, H., and Gong, G.: Disrupted white matter connectivity underlying developmental dyslexia: A machine learning approach. Human Brain Mapping, 37 (4), 1443–1458 (2016).
17. Uhrig, R.E.: Introduction to artificial neural networks. In: Proceedings of IECON '95 – 21st Annual Conference on IEEE Industrial Electronics, vol. 1, pp. 33–37 (1995).
18. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15 (1), 1929–1958 (2014).
19. García Chimeno, Y., García Zapirain, B., Saralegui Prieto, I., and Fernandez-Ruanova, B.: Automatic classification of dyslexic children by applying machine learning to fMRI images. Biomedical Materials and Engineering, 24 (6), 2995–3002. doi:10.3233/BME-141120 (2014).
20. Jolliffe, I.T. and Cadima, J.: Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A, 374 (2065):20150202 (2016).
21. McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., and Sejnowski, T.J.: Analysis of fMRI data by blind separation into independent spatial components. Human Brain Mapping, 6 (3), 160–188 (1998).
22. Silva, L.T.D., Esper, N.B., Ruiz, D.D., Meneguzzi, F., and Buchweitz, A.: Visual Explanation for Identification of the Brain Bases for Developmental Dyslexia on fMRI Data. Frontiers in Computational Neuroscience, 15, 594659 (2021).
23. Taran, N., Gatenyo, R., Hadjadj, E., Farah, R., and Horowitz-Kraus, T.: Distinct connectivity patterns between perception and attention-related brain networks characterize dyslexia: Machine learning applied to resting-state fMRI. Cortex, 181, 216–232 (2024).
24. OpenNeuro, https://openneuro.org/, last accessed 2024/10/23
25. Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., et al.: The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular Psychiatry, 19 (6), 659–667 (2014).

26. Essen, D.C.V., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., and the WU-Minn HCP Consortium.: The WU-Minn Human Connectome Project: An overview. NeuroImage, 80, 62–79 (2013).
27. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D.: Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017).
28. Poldrack, R.A., Mumford, J.A., and Nichols, T.E.: Handbook of fMRI Data Analysis. Cambridge University Press, Cambridge (2011).
29. Friston, K.J., Ashburner, J., Kiebel, S.J., Nichols, T.E., and Penny, W.D.: Statistical parametric mapping. Elsevier Academic Press, New York (2007).
30. Vayena, E., Blasimme, A., and Cohen, I.G.: Machine learning in medicine: Addressing ethical challenges. PLOS Medicine, 15 (11), e1002689 (2018).
31. Sui, J., Adali, T., Yu, Q., Chen, J., and Calhoun, V.D.: A review of multivariate methods for multimodal fusion of brain imaging data. Journal of Neuroscience Methods, 204 (1), 68–81 (2012).

HOW TO CITE: Adarsh Pradhan, Shikhar Kumar Sarma, Mirzanur Rahman, A Review of Machine Learning-Based Dyslexia Detection Using fMRI Data, Int. J. Sci. R. Tech., 2026, 3 (11), 870-876. https://doi.org/10.5281/zenodo.24300321