Toward Personalized Dyslexia Classification via Dynamic Functional Connectivity and Explainable AI

Blessy Thomas, Moutaz Saleh, Somaya Al Maadeed, Younes Akbari, and Jayakanth Kunhoth

Department of Computer Science and Engineering, Qatar University, Doha, Qatar
blessythomas1296@gmail.com
{moutaz.saleh, s_alali, younes.akbari, jayakanth.kunhoth}@qu.edu.qa

Abstract. Dyslexia is a neurological disorder associated with difficulties in reading and decoding. Neuroimaging studies have shown that individuals with dyslexia exhibit altered brain connectivity during reading-related tasks. However, existing works have not studied the temporal dynamics of neural interactions during reading tasks. This study integrates Dynamic Functional Connectivity (DFC) analysis with Explainable Artificial Intelligence (XAI) to investigate dyslexia-related alterations in brain networks using task-based fMRI (functional Magnetic Resonance Imaging) data. Sliding window-based ROI-to-ROI connectivity features were extracted from 200 cortical parcels for classification. SHapley Additive exPlanations (SHAP) were employed to identify the most informative connectivity patterns. The proposed framework achieved a classification accuracy of 0.758 ± 0.091 using XGBoost. A gender-stratified evaluation further revealed differences in classification performance, with RBF-SVM achieving 0.819 ± 0.053 for males and XGBoost achieving 0.840 ± 0.150 for females. SHAP analysis identified connectivity involving left-hemisphere language-related and parietal regions, and their connectivity with default mode, salience, and cognitive control networks, as key discriminative features for dyslexia classification. These findings support interpretable dyslexia screening and highlight the importance of gender-specific evaluation in neuroimaging-based dyslexia classification.

Keywords: Dyslexia · Functional Magnetic Resonance Imaging (fMRI) · Dynamic Functional Connectivity (DFC)· Region of Interest (ROI)· Explainable AI (XAI).

1 Introduction

Dyslexia is a learning disorder characterized by reading difficulties that can lead to long-term academic, emotional, and social consequences [8]. It affects up to 17.5% of children and is associated with alterations in brain connectivity during reading and language-related tasks [6]. Understanding these neural differences

is critical for supporting educators and clinicians in providing appropriate interventions. Functional magnetic resonance imaging (fMRI) enables the investigation by analyzing brain activity and connectivity patterns during reading tasks [15]. However, most studies rely on static functional connectivity, which captures average connectivity patterns but does not represent the temporal dynamics of neural interactions that occur during reading [10, 1, 2]. Dynamic Functional Connectivity (DFC) addresses this limitation by modeling time-varying changes in brain connectivity during reading.
Machine learning models applied to dyslexia classification operate as black boxes, producing predictions without interpretable justification. In a clinical and educational screening context, interpretability is not optional, practitioners, educators, and clinicians need to understand which neural patterns drive a classification decision in order to trust and act on it. Explainable artificial intelligence (XAI) methods bridge this gap by identifying the important connectivity features for classification decisions. Integrating DFC with XAI enables the identification of which brain connectivities are most discriminative for dyslexia, providing deeper insight into the neural basis of reading difficulties.

Another limitation of existing studies is the assumption that all participants exhibit similar connectivity patterns. Prior research has reported gender-related differences in brain connectivity [14, 17], suggesting that dyslexia-related neural signatures may vary between male and female participants. Consequently, gender-stratified evaluation may provide a more informative characterization of dyslexia-related connectivity patterns than a single population-level model.

In this work, we propose a framework that integrates Dynamic Functional Connectivity analysis with SHAP (SHapley Additive exPlanations) to investigate dyslexia-related alterations in brain networks using task-based fMRI data. Time-varying ROI-to-ROI connectivity features are extracted and used for classification, while a gender-stratified evaluation examines whether classification performance and connectivity signatures differ across genders. By identifying the most discriminative ROI-to-ROI interactions and their temporal dynamics, the proposed framework provides an interpretable characterization of dyslexia-related brain connectivity.

The major contributions of this work are as follows:

- We extract time-varying ROI-to-ROI connectivity features from task-based fMRI data and demonstrate their effectiveness for dyslexia classification.
- We introduce a gender-stratified evaluation framework that reveals differences in classification performance between male and female participants.
- We employ SHAP-based explainability to identify neurobiologically meaningful connectivity patterns associated with dyslexia, providing interpretable insights into classification decisions.

2 Related Work

Dyslexia classification has been extensively studied using both traditional diagnostic methods and machine learning approaches. Existing machine learn-

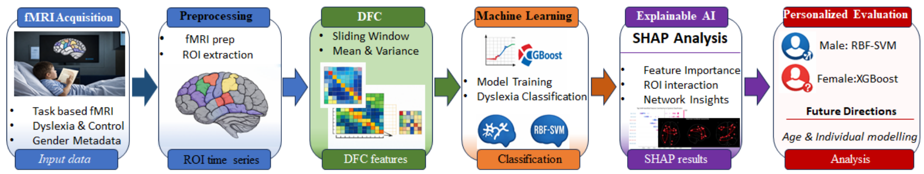


Fig. 1: Overview of the proposed DFC-based dyslexia classification and explainability framework.

ing–based studies on dyslexia classification employ diverse data modalities, including handwriting, eye-tracking, EEG, MRI, and fMRI [10],[2],[3],[9] to capture behavioral, neural, and visual processing differences associated with dyslexia. In this section, we focus on fMRI-based dyslexia classification studies.

Several studies have applied deep learning techniques to fMRI-based dyslexia classification. CNN-based approaches typically treat fMRI volumes as images by converting 3D scans into 2D slices and extracting deep features using architectures such as Bayesian-optimized CNNs, ShuffleNet V2, and MobileNet V3 [10, 1, 2]. Multimodal frameworks combining fMRI with MRI, diffusion MRI, or EEG have also been explored, utilizing deep learning or ROI-based activation features for dyslexia identification [2, 5].

Task-based and resting-state fMRI have also been explored for dyslexia classification. Deep learning approaches, including 3D CNNs and CNN–LSTM architectures, have been used to learn spatial and temporal patterns from reading-related brain activity, while multimodal frameworks have combined structural MRI and fMRI to improve classification performance [16, 7]. Functional connectivity-based studies have further demonstrated the importance of interactions among large-scale brain networks, particularly attention-related systems, for distinguishing individuals with dyslexia from controls [13].

Although substantial research has been conducted on dyslexia detection, existing studies lack interpretability of brain connectivity patterns during reading tasks, and gender-specific evaluation remains underexplored in fMRI-based classification. These limitations motivate the proposed framework, which integrates XAI to interpret time-varying neural interactions during reading, alongside a gender-stratified evaluation to examine differences in dyslexia-related brain connectivity between male and female participants.

3 Methodology

3.1 Dataset Description and Preprocessing

The dataset [4] includes 3T fMRI data from 58 children (34 males, 24 females): 22 typical readers, 16 with isolated spelling deficits (SpD), and 20 with dyslexia. Participants read 60 words and 60 pseudohomophones aloud in an event-related design across three runs. For binary classification, typical readers and SpD children were combined into a single control class. Preprocessing was performed using fMRIPrep (version 23.2.1) with spatial normalization to MNI space. Figure 1 shows the overview of the proposed framework.

3.2 Feature Extraction

DFC analysis was performed using the Schaefer 200 atlas [12] to define 200 cortical ROIs. For each ROI, voxel-wise BOLD (Blood-Oxygen-Level Dependent) signals were averaged and z-score normalized. Time-varying connectivity was captured using a sliding-window approach (window = 30 TRs, step = 1 TR), with pairwise Pearson correlations computed within each window:

$$\mathrm{FC}_{ij}^{(w)} = \mathrm{corr}\left(\hat{x}_i^{(w)}, \hat{x}_j^{(w)}\right) \tag{1}$$

, where $\hat{x}_i^{(w)}$ and $\hat{x}_j^{(w)}$ denote the windowed, detrended, and z-score normalized time series of ROIs i and j, respectively. For each ROI pair (i, j), the temporal mean and variance of the connectivity strength across all sliding windows were calculated to characterize stable and dynamic connectivity patterns:

$$\mu_{ij} = \frac{1}{W}\sum_{w=1}^{W} \mathrm{FC}_{ij}^{(w)}, \quad \sigma_{ij}^2 = \frac{1}{W}\sum_{w=1}^{W}\left(\mathrm{FC}_{ij}^{(w)} - \mu_{ij}\right)^2 \tag{2}$$

, where W denotes the total number of sliding windows. μ_{ij} is the temporal mean, and σ_{ij}^2 is the temporal variance.

To reduce redundancy, only the upper triangular elements of the connectivity matrices (excluding the diagonal) were retained. The final feature vector for each subject was constructed by concatenating the mean and variance values of all unique ROI–ROI pairs and was used as input for subsequent machine learning and explainability analyses.

$$\mathbf{DFC} = \left[\{\mu_{ij}\}_{i<j}, \ \{\sigma_{ij}^2\}_{i<j}\right] \tag{3}$$

3.3 Classification and Explainability

DFC features were used to classify subjects into dyslexia and control groups. A 5-fold stratified cross-validation was employed. Within each fold, features were standardized and dimensionality was reduced using principal component analysis (PCA) while retaining 99% of the variance. XGBoost was selected as the final classifier because it achieved the highest classification performance among all the evaluated machine learning models. It is modeled as an additive ensemble of decision trees:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\mathbf{x}_i) \tag{4}$$

, where K denotes the number of trees and f_k represents the k-th decision tree in the ensemble. SHAP was applied to quantify feature contributions to model predictions [11]:

$$f(x) = \phi_0 + \sum_{j=1}^{d} \phi_j \tag{5}$$

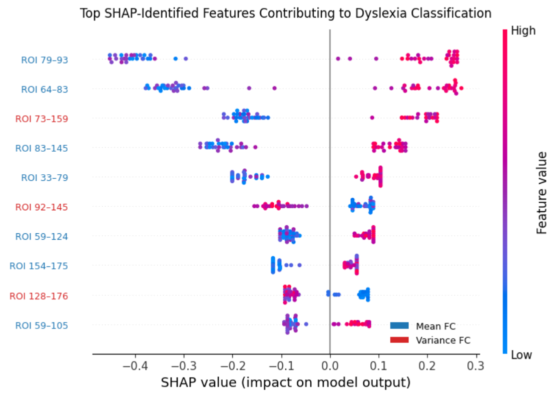


Fig. 2: SHAP-important ROI–ROI connectivity features for dyslexia classification

, where ϕ_0 is the baseline expected output and ϕ_j is the contribution of feature j. Gender-stratified models were further trained separately for male and female participants to examine gender-specific classification performance and connectivity patterns.

4 Results and Discussion

A systematic evaluation was conducted to assess the performance of the proposed framework for dyslexia classification. All experiments were implemented in Python using traditional machine learning models. XGBoost and RBF-SVM classifiers were implemented using the Scikit-learn library, and SHAP was employed for feature importance visualization and model interpretation. A stratified 5-fold cross-validation procedure was employed, with predictions aggregated across folds such that each subject contributed exactly one prediction.

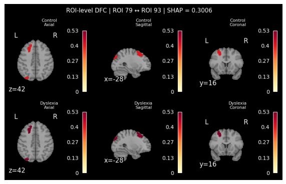
Table 1: Model-wise Classification Performance

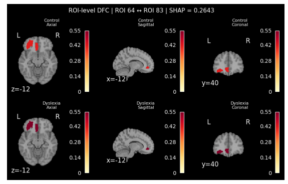
Model	**Accuracy (Mean ± Std)**
Decision Tree	0.5394 ± 0.1842
RBF-SVM	0.6667 ± 0.0332
Logistic Regression	0.6818 ± 0.1323
KNN	0.6848 ± 0.0680
XGBoost	$\mathbf{0.7581 \pm 0.0913}$

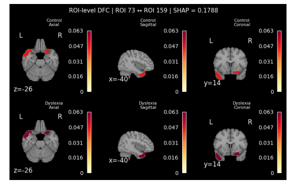
Table 1 presents the classification performance across all evaluated models. XGBoost achieved the highest mean accuracy of 0.758 ± 0.091, demonstrating superior performance compared to RBF-SVM (0.666 ± 0.033), Logistic Regression (0.681 ± 0.132), and KNN (0.684 ± 0.068), while Decision Tree exhibited the lowest performance (0.539 ± 0.184). These results indicate that ensemble-based

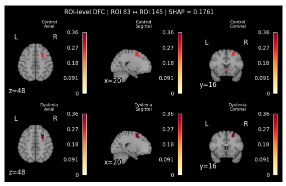
methods are more effective in capturing discriminative DFC patterns associated with dyslexia, likely due to their ability to model non-linear interactions among time-varying connectivity features. Figure 2 illustrates the most influential ROI–ROI DFC features identified through SHAP analysis, ranked by their mean absolute SHAP values. Both mean and variance measures of ROI–ROI connectivity are identified as discriminative, where the mean captures sustained coupling strength and the variance reflects temporal instability of neural connectivity.
The most influential feature involves functional connectivity between the left temporal default mode region (ROI 79: Default_Temp_5) and the left medial prefrontal default mode region (ROI 93: Default_PFC_10). The left temporal default mode region is associated with semantic processing, language comprehension, and conceptual knowledge, whereas the left medial prefrontal default mode region is involved in self-referential processing, semantic memory retrieval, and higher-order cognitive functions. Both regions are components of the Default Mode Network (DMN), which has been reported to exhibit altered functional connectivity in individuals with dyslexia, reflecting atypical integration of language-related and higher-order cognitive processes. Another important feature involves functional connectivity between the left parietal control region (ROI 64: Cont_Par_3) and the left parietal default mode region (ROI 83: Default_Par_4). The left parietal control region contributes to executive control, attention allocation, and working memory, while the left parietal default mode region is associated with internally directed cognition and memory-related processing. Altered functional connectivity between the executive control and default mode networks has been associated with the increased cognitive effort and compensatory processing observed during reading in individuals with dyslexia. In addition, the variance of functional connectivity between the left cingulate control region (ROI 73: Cont_Cing_1) and the right medial salience/ventral attention region (ROI 159: SalVentAttn_Med_3) emerged as an important discriminative feature. The left cingulate control region is involved in cognitive control, conflict monitoring, and task regulation, whereas the right medial salience/ventral attention region supports salience detection, attentional reorientation, and dynamic switching between large-scale brain networks. Previous neuroimaging studies have reported atypical functional connectivity within the cognitive control and salience networks in dyslexia, suggesting altered attentional regulation and executive control during language and reading tasks. Figure 3 illustrates group-wise ROI–ROI DFC differences between dyslexic and control participants. In most cases, individuals with dyslexia exhibit higher mean connectivity and greater temporal variability compared to controls, reflecting widespread disruption across default mode, control, and attentional networks characterized by increased and poorly regulated connectivity particularly within left-hemisphere language networks.

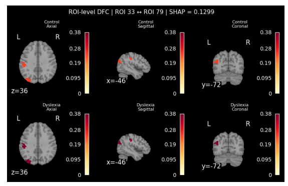
Table 2 reports the gender-wise classification performance. The gender-wise analysis shows that RBF-SVM achieves optimal performance for male participants (0.819 ± 0.053), while XGBoost performs best for females (0.840 ± 0.150),

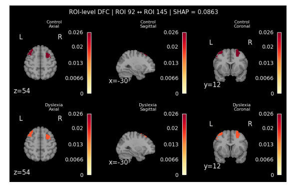

(a) ROI 79–ROI 93

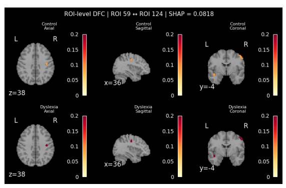

(b) ROI 64–ROI 83

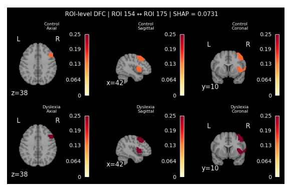

(c) ROI 73–ROI 159


(d) ROI 83–ROI 145


(e) ROI 33–ROI 79


(f) ROI 92–ROI 145


(g) ROI 59–ROI 124


(h) ROI 154–ROI 175

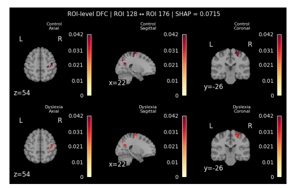

(i) ROI 128–ROI 176

Fig. 3: Group-wise differences in ROI–ROI connectivity between subjects with dyslexia and typically developing controls

demonstrating that no single classifier uniformly captures dyslexia-related connectivity patterns across genders. Furthermore, the gender-specific connectivity pattern analysis shown in Figure 4 confirms that the underlying neural signatures associated with dyslexia differ between male and female participants. This motivates a broader personalization hierarchy for dyslexia classification. At the first level, demographic-level personalization tailors feature extraction and classifier selection to gender-based subgroups, as demonstrated in this work. At the second level, age-stratified personalization would extend this framework to developmental subgroups. Age-wise connectivity patterns are visualized in Figure 5, a full age-stratified classification analysis is not pursued here due to the limited number

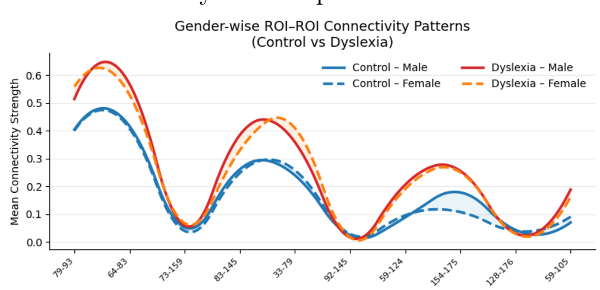
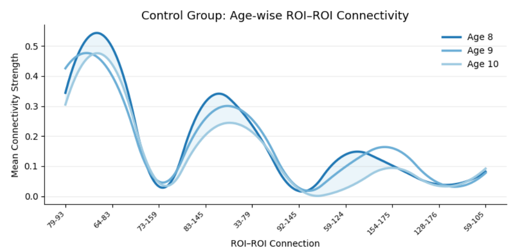
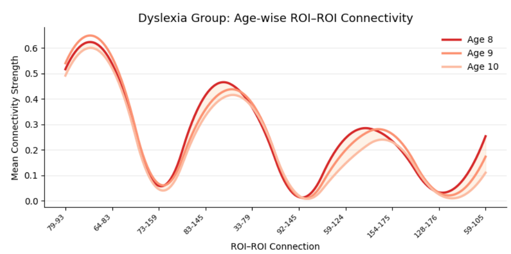


Fig. 4: Gender-wise differences in ROI-ROI connectivity patterns between control and subjects with dyslexia



(a) Control group



(b) Dyslexia group

Fig. 5: Age-wise differences in ROI–ROI Connectivity for Control and Dyslexia groups

of subjects per age group in the present dataset At the third level, individual-level personalization would involve dedicated pipelines per participant, requiring substantially larger longitudinal datasets than currently available, and is identified as an important future direction.

Table 2: Gender-wise Classification Performance

Gender	Model	Accuracy (Mean ± Std)
Male	**RBF-SVM**	$\mathbf{0.819 \pm 0.053}$
Male	XGBoost	0.690 ± 0.051
Female	RBF-SVM	0.750 ± 0.150
Female	**XGBoost**	$\mathbf{0.840 \pm 0.150}$

5 Conclusion

This work analyzes dynamic functional connectivity for dyslexia–control classification using machine learning with SHAP-based interpretability. The results demonstrate that classification is primarily influenced by functional connectivity within left-hemisphere language-related and parietal regions, and their connectivity with the default mode, executive control, and salience/attention networks. SHAP analysis identified both the mean and variance of dynamic functional connectivity as important discriminative features, suggesting that both sustained connectivity patterns and their temporal variability contribute to distinguishing individuals with dyslexia from typical controls. Beyond overall classification, this study advances personalized dyslexia modeling through gender-stratified evaluation, showing that different classifiers and distinct connectivity patterns for males and females underscore the value of demographic-aware approaches as a principled first step toward broader personalization frameworks.

6 Acknowledgment

This publication was supported by Qatar University High Impact Grant QUHI-CENG-25/26-665. The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of Qatar University.

References

1. Alkhurayyif, Y., Sait, A.R.W.: Developing an image-based dyslexia detection model using the deep learning technique. Journal of Disability Research **2**(4), 79–88 (2023)
2. Alkhurayyif, Y., Sait, A.R.W.: Deep learning-driven dyslexia detection model using multi-modality data. PeerJ Computer Science **10**, e2077 (2024)
3. Alqahtani, N.D., Alzahrani, B., Ramzan, M.S.: Detection of dyslexia through images of handwriting using hybrid ai approach. International Journal of Advanced Computer Science and Applications **14**(10) (2023)
4. Banfi, C., Koschutnig, K., Moll, K., Schulte-Koerne, G., Fink, A., Landerl, K.: "reading-related functional activity in children with isolated spelling deficits and dyslexia" (2022). https://doi.org/doi:10.18112/openneuro.ds003126.v1.3.1
5. Cristina, C., Giulia, Z., Marta, G., Marco, T.: Multimodal investigation of the neurocognitive deficits underlying dyslexia in adulthood. bioRxiv pp. 2024–11 (2024)
6. Habib, M., Giraud, K.: Dyslexia. Handbook of clinical neurology **111**, 229–235 (2013)
7. LR, H., Sadasivam, G.S.: Multimodal screening for dyslexia using anatomical and functional mri data. Journal of Computational Methods in Science and Engineering **22**(4), 1105–1116 (2022)
8. Nemmi, F., Cignetti, F., Vaugoyeau, M., Assaiante, C., Chaix, Y., Péran, P.: Developmental dyslexia, developmental coordination disorder and comorbidity discrimination using multimodal structural and functional neuroimaging. Cortex **160**, 43–54 (2023)
9. Prabha, A.J., Bhargavi, R.: Predictive model for dyslexia from fixations and saccadic eye movement events. Computer Methods and Programs in Biomedicine **195**, 105538 (2020)
10. Pradhan, A., Rahman, M., Sarma, S.K., Kalita, A., Ali, A., Ghosh, P.: Classification of dyslexia using bayesian optimized cnn on plane-wise separated fmri images with visual interpretation. IETE Journal of Research **71**(7), 2475–2488 (2025)
11. Salih, A.M., Raisi-Estabragh, Z., Galazzo, I.B., Radeva, P., Petersen, S.E., Lekadir, K., Menegaz, G.: A perspective on explainable artificial intelligence methods: Shap and lime. Advanced Intelligent Systems **7**(1), 2400304 (2025)
12. Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.: Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. Cerebral cortex **28**(9), 3095–3114 (2018)

13. Taran, N., Gatenyo, R., Hadjadj, E., Farah, R., 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)
14. Tomasi, D., Volkow, N.D.: Gender differences in brain functional connectivity density. Human brain mapping **33**(4), 849–860 (2012)
15. Usman, O.L., Muniyandi, R.C.: Cryptodl: Predicting dyslexia biomarkers from encrypted neuroimaging dataset using energy-efficient residue number system and deep convolutional neural network. Symmetry **12**(5), 836 (2020)
16. Zahia, S., Garcia-Zapirain, B., Saralegui, I., Fernandez-Ruanova, B.: Dyslexia detection using 3d convolutional neural networks and functional magnetic resonance imaging. Computer methods and programs in biomedicine **197**, 105726 (2020)
17. Zhang, S., Li, M., Chen, A., Jiang, L., Yu, H., Chen, Y.: Sex differences and age-related changes of large-scale brain networks. BMC Medical Imaging **25**(1), 271 (2025)