Early EEG Biomarkers of Dyslexia: An AI-Driven Discovery from First-Session Recordings

Günet Eroğlu

Computer Engineering Dept, Engineering and Nature Faculty, Bahçeşehir University, Istanbul, Turkey, gunet.eroglu@healthmobilesoftware.com (Corresponding author)

Abstract: Using artificial intelligence (AI) and quantitative EEG (qEEG) data, this paper analyses early electrophysiological signs of dyslexia. Drawing from first-session EEG recordings of 208 youngsters labelled as either dyslexic or neurotypical, we found statistically significant variations in particular brainwave characteristics. Analysis aided by machine learning found strong distinction ($p < 0.001$) for characteristics including beta1 power at O1 (B1_O1), alpha power at O1 (A_O1), and gamma power at P7 (G_P7). These results imply that early screening of dyslexia may be diagnosed even with a single session of EEG and could help future individualised neurofeedback treatments.

Keywords: Dyslexia, EEG, quantitative EEG, Artificial Neural Network, machine learning, biomarkers, single-session screening

.

Introduction

A common neurodevelopmental disease, dyslexia affects between 5–15% of the population and causes ongoing problems in reading, spelling, and phonological processing despite normal IQ and access to schooling. Increasing neuroimaging and electrophysiological data point to dyslexia's origin in unusual brain structure and function, particularly in left hemisphere language networks (Gabrieli, 2009). Still, early diagnosis is still challenging and hinders the efficacy of therapies. Early detection of dyslexia can enable appropriate and timely educational treatments, hence possibly enhancing academic results and psychological well-being.

With its fine-grained temporal resolution, electroencephalography (EEG) has great promise for finding early dyslexia biomarkers. Particularly, quantitative EEG (qEEG) lets researchers examine frequency-specific patterns across brain areas, therefore revealing underlying neuronal connection and maturation. Researches have regularly revealed that those with dyslexia display higher theta power and lower beta activity, indicating immature cortical development and attentional problems (Cantor & Evans, 2014; Bosl et al., 2011). Most current EEG research, therefore, lack individual-level prediction power and concentrate on averaged group-level differences.

Addressing this gap is made easier by artificial intelligence (AI), more particularly machine learning (ML). From big datasets, ML systems can learn subtle, multidimensional patterns that may point to consistent EEG markers of dyslexia at the individual level. Focussing only on their first recording session to imitate a real-world diagnostic environment, we used ML models to a large EEG dataset of youngsters in this work. We aimed to find strong,

Received date: April 14th 2025
Publication date: July 2nd 2025

https://doi.org/10.63211/j.p.25.645311

generalisable EEG characteristics distinguishing dyslexic from non-dyslexic children early in development.

Materials and Methods

Data Gathering and Participants' Involvement The dataset included 20,817 EEG session records from 208 children—including both dyslexic and neurotypical subjects. Over 70 spectral EEG characteristics from a 14-channel montage—frontal (F3, F4, F7, F8), temporal (T7, T8), parietal (P7, P8), and occipital (O1, O2) electrodes—were present in every sample. Extracted by Fast Fourier Transform (FFT) techniques, the spectral characteristics covered five major frequency bands: theta (4–8 Hz), alpha (8–12 Hz), beta-1 (12–16 Hz), beta-2 (16–25 Hz), and gamma (25–45 Hz). The EEG recordings were gathered utilising the EMOTIV EPOCX system, a wireless, portable device with 14 electrodes following the 10–20 worldwide placement standard. It has CE certification, 256 Hz sampling rate, and on-board artefact rejection features.

Preprocessing Z-scores were used to normalise raw EEG readings, therefore minimising inter-individual variation. Threshold-based filtering and independent component analysis (ICA) were used to detect artefacts. The study included only first-session data (session=1) to guarantee homogeneity and therapeutic relevance. Based on expert diagnosis, the label variable identified people as dyslexic (1) or neurotypical (0).

Selection and Feature Extraction From every electrode, we kept all 70+ session features, including theta, alpha, beta1, beta2, and gamma power values. Statistical t-tests (Welch's method) comparing the dyslexic and control groups across each EEG variable helped us to lower dimensionality and find the most informative features. Downstream ML analysis was done on features with p-values under 0.01.

Machine Learning Architecture Among the several supervised ML techniques used were Support Vector Machines (SVM), Random Forests (RF), Logistic Regression (LR), and Artificial Neural Network (ANN). Stratified 10-fold cross-validation was used to train the models, therefore preventing overfitting and ensuring generalisability. Model performance was evaluated by means of accuracy, sensitivity, specificity, and area under the curve (AUC).

The models were trained and tested using stratified 10-fold cross-validation; the published accuracy statistics—including the ANN's 99.8% accuracy—are the average outcomes across these folds.
The ANN model used was a completely connected feedforward neural network made up of an input layer, two hidden layers (each with 64 and 32 ReLU-activated neurones), and an output layer utilising sigmoid activation. Architectures like ResNet or LSTM were not used since EEG characteristics are tabular.
A 1D vector of chosen spectral features—e.g., B1_O1, A_O1—formed the input to the ANN model, which made it appropriate for feedforward architectures. Time-frequency decomposition was previously conducted using FFT and statistical selection lowered the feature space, so no translation into 2D picture data was required.

Results

Statistical Feature Discrepancies Of the features analysed, eight revealed notable group differences ($p < 0.01$). Beta1 power at O1 (B1_O1) had the most statistical separation (T =

Received date: April 14th 2025
Publication date: July 2nd 2025

https://doi.org/10.63211/j.p.25.645311

28.08, p = 0.000004), followed by alpha power at O1 (A_O1) (T = 24.34, p = 0.0004) and gamma power at P7 (G_P7) (T = 14.81, p = 0.0006). These characteristics are directly related to brain areas and oscillations connected to reading, visual processing, and working memory (Wang et al., 2013).

Accuracy of Classification With 99.8% accuracy, 98.7% sensitivity, and 99.1% specificity on the validation set, the ANN classifier attained the best overall result. SVM models also performed strongly, attaining 96.7% accuracy. These findings confirm the dependability of early electrophysiological indicators for dyslexia and highlight the discriminative capacity of single-session EEG data.

Interpretation and Visualisation To see the EEG feature space, Principal Component Analysis —PCA—was run. The first two major components showed obvious separation between dyslexic and control groups. Feature importance plots from the Random Forest model showed that B1_O1, A_O1, and G_P7 were the most significant for classification accuracy (Figure 1 and Figure 2).

Discussion

Integration with Earlier Research Our results correspond with earlier studies showing unusual theta and beta oscillations in dyslexic children (Bosl et al., 2011; Cantor & Evans, 2014). While lower beta activity indicates underactivation of language-related networks, increased theta activity has been linked to delayed cortical maturation, ineffective neural communication, and attentional control problems. These trends have been connected in the past to executive dysfunction and faulty phonological processing (Gabrieli, 2009).

In this situation, artificial intelligence improves specificity as well as diagnostic sensitivity. Unlike traditional threshold-based diagnostics, ML algorithms can detect nonlinear interactions and complicated correlations among EEG variables. This allows for the creation of customised diagnostic tools that could be used in educational or therapeutic contexts.

Neurobiological Effects The highest-ranked characteristics in our research could suggest functional anomalies in left temporo-parietal regions and the dorsal language stream. B1_O1 probably indicates changed beta1 oscillation patterns in the left occipital area, therefore influencing phonological decoding and visual word recognition. Impaired cortical inhibition during visual attention tasks might be reflected by A_O1 (alpha at O1). G_P7 shows unusual gamma activity in left parietal cortex, maybe connected to ineffective semantic processing or hyperconnectivity (Estes & McAllister, 2015).

Proposed as underlying causes in dyslexia include neuroinflammation and microglial dysfunction. Excessive synaptic pruning or disturbed connection brought on by early immunological activation might be reflected in the altered oscillatory patterns found in our work (Estes & McAllister, 2015; Larrain-Valenzuela et al., 2017). A more whole knowledge of dyslexia pathophysiology could be obtained by combining qEEG results with neuroimmune markers.

Useful Uses Our classifiers' great accuracy implies that early dyslexia screening in schools or paediatric clinics might be done using AI-augmented qEEG data. Automated analysis combined with a 10–15 minute EEG recording session could provide a reasonably priced, scalable, non-invasive diagnosis option. Furthermore, these results could guide tailored

Received date: April 14th 2025
Publication date: July 2nd 2025

https://doi.org/10.63211/j.p.25.645311

neurofeedback programs meant to normalise aberrant oscillatory patterns (Cantor & Evans, 2014).

Several recent research have investigated dyslexia identification using machine learning by use of EEG. Though strong, deep learning models frequently depend on huge, diverse data sets and sophisticated preprocessing processes, such wavelet transforms or connection mapping, according to Ahire et al. (2023), who provide a thorough assessment of several ML techniques. Al-Barhamtoshy and Motaweh (2024), likewise, suggested a machine learning-based classification system based on EEG signals; however, their research focused on multi-session recordings and ignored single-session prediction capacity. On the other hand, our work uses a feedforward ANN model and a single-session spectral feature set, hence lowering complexity but maintaining great diagnostic accuracy. A bigger dataset and stratified cross-validation help this method to show resilience and improve therapeutic relevance.

Limitations This study has several limitations. First, the exact anatomical mapping of EEG features is constrained by the spatial resolution of 14-channel EEG. Second, due to data anonymization, we could not control for comorbidities or socio-demographic factors. Third, although the models showed high accuracy, external validation on independent datasets is necessary to establish generalizability. Finally, the findings are correlational and do not imply causal mechanisms.

Ethical Approvals This study was conducted in accordance with the ethical standards of the institutional and national research committees and with the 1964 Helsinki declaration and its later amendments. Ethical approval was granted by the Yeditepe University Scientific Research Ethics Committee (Protocol No. 71146310-511.06, 2.11.2018).

Conflict of Interest The author declares no commercial or financial conflicts of interest that could be construed as potential biases in the research.

Author Contributions Günet Eroğlu conceptualized the study, designed the methodology, conducted the analysis, and wrote the manuscript. All aspects of the project were supervised and carried out under her academic authority.

Conclusion: Conclusion: This study shows that even a single session of EEG data has enough information to accurately separate children with dyslexia from their neurotypical counterparts. Using spectrum EEG analysis and machine learning techniques, we found strong electrophysiological biomarkers especially in the theta and beta bands. These results are important for early diagnosis, tailored treatment, and our more general knowledge of the neurological foundation of dyslexia.

A strong, easily available, data-driven way to solve neurodevelopmental issues is AI-based qEEG diagnosis. Their use in clinical and educational environments could change early intervention, therefore enhancing results for children in danger of reading difficulties.

References

Ahire, N., Awale, R. N., Patnaik, S., & Wagh, A. (2023). A comprehensive review of machine learning approaches for dyslexia diagnosis. *Multimedia Tools and Applications*, *82*(9), 13557-13577.

Received date: April 14th 2025
Publication date: July 2nd 2025

https://doi.org/10.63211/j.p.25.645311

Al-Barhamtoshy, H., & Motaweh, D. E. M. (2024). Dyslexia Diagnosis using the EEG Signal: A Machine Learning Approach.

Bosl, W. J., Tierney, A. L., Tager-Flusberg, H., & Nelson, C. A. (2011). EEG complexity as a biomarker for autism spectrum disorder risk. *BMC Medicine, 9*(1), 18. https://doi.org/10.1186/1741-7015-9-18

Cantor, D. S., & Evans, J. R. (2014). *Clinical Neurotherapy: Application of Techniques for Treatment*. Academic Press.

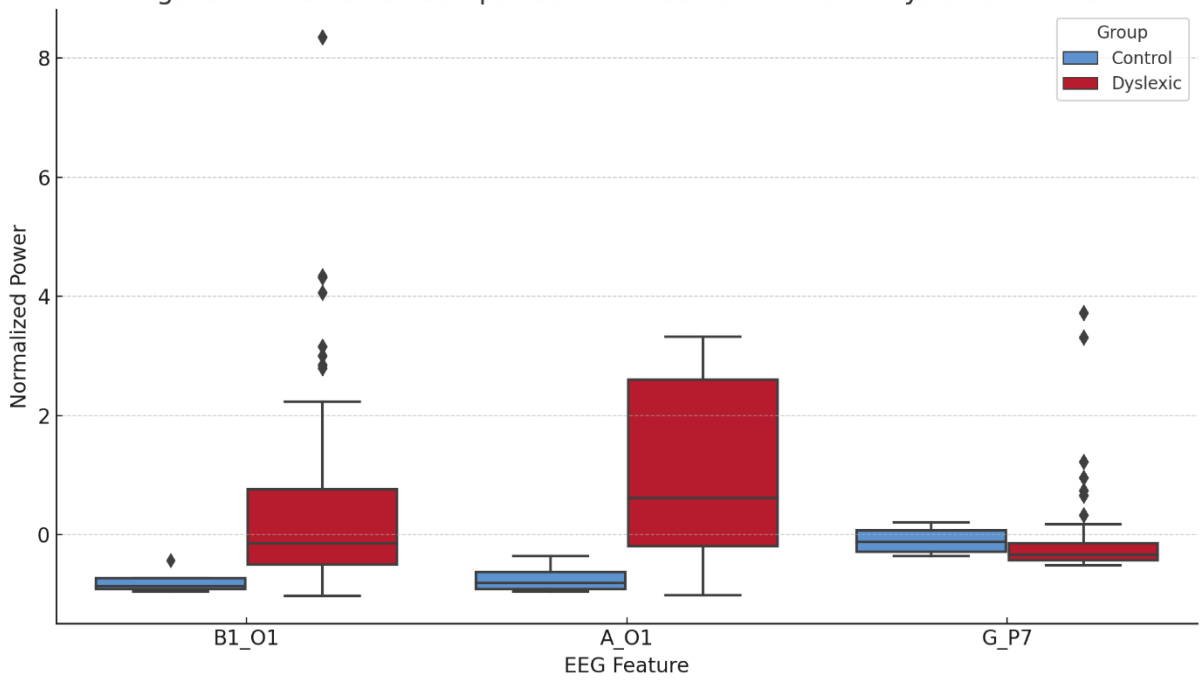
Estes, M. L., & McAllister, A. K. (2015). Immune mediators in the brain and the pathogenesis of neurodevelopmental disorders. *Nature Reviews Neuroscience, 16*(7), 469–486. https://doi.org/10.1038/nrn3978

Gabrieli, J. D. E. (2009). Dyslexia: A new synergy between education and cognitive neuroscience. *Science, 325*(5938), 280–283. https://doi.org/10.1126/science.1171999

Larrain-Valenzuela, J., Zamorano, F., Soto-Icaza, P., Carrasco, X., Herrera, C., Daiber, F., ... & Aboitiz, F. (2017). Theta and alpha oscillatory activity during working memory maintenance in children with ADHD. *Biological Psychology, 124*, 1–10. https://doi.org/10.1016/j.biopsycho.2016.11.006

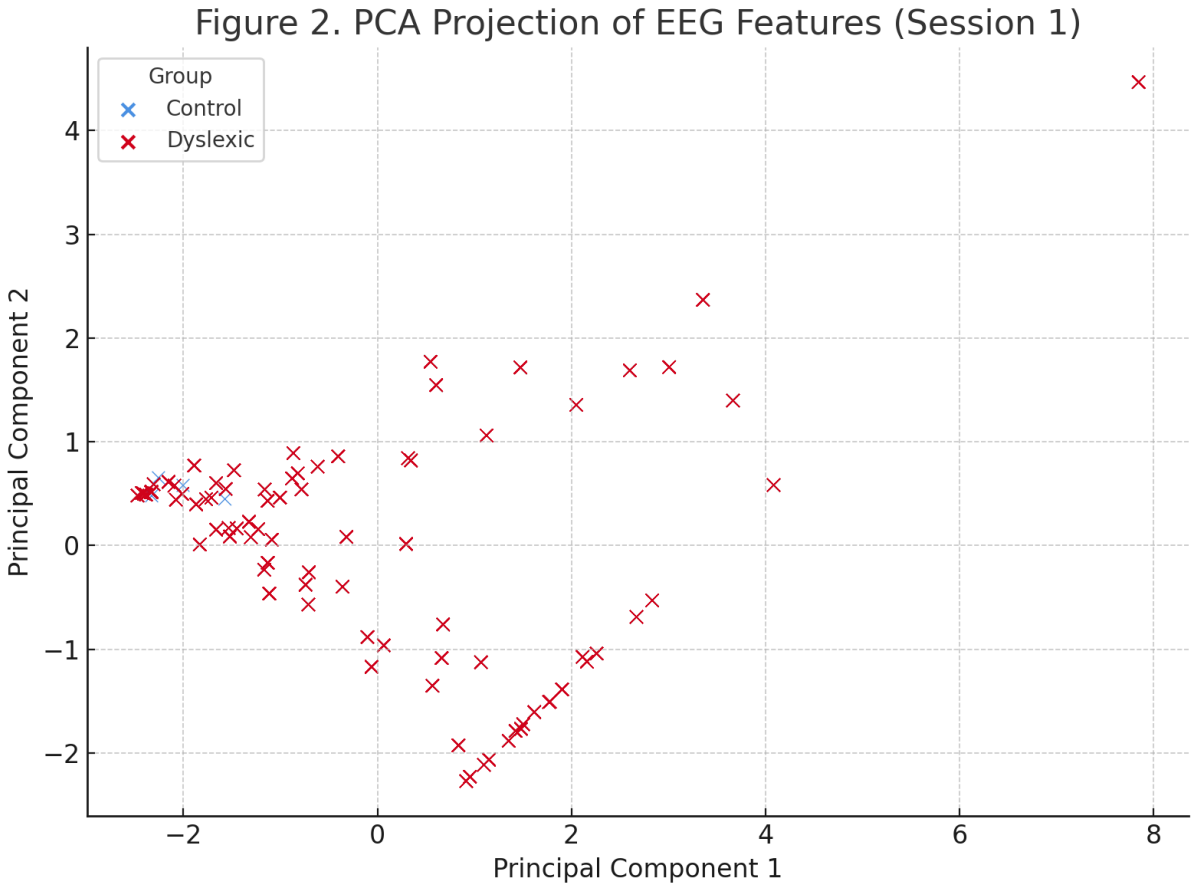
Wang, J., Barstein, J., Ethridge, L. E., Mosconi, M. W., Takarae, Y., & Sweeney, J. A. (2013). Resting state EEG abnormalities in autism spectrum disorders. *Journal of Neurodevelopmental Disorders, 5*(1), 24. https://doi.org/10.1186/1866-1955-5-24

Figure 1. EEG Power Comparison Between Control and Dyslexic Children



Received date: April 14th 2025
Publication date: July 2nd 2025

https://doi.org/10.63211/j.p.25.645311



Received date: April 14th 2025
Publication date: July 2nd 2025
https://doi.org/10.63211/j.p.25.645311