Transactions on Machine Intelligence 8(2) (2025) 120–127

ISSN Online: 2821-1693

Transactions on Machine Intelligence

Journal Homepage: https://tmachineintelligence.ir/

Leveraging Bagging Ensemble Architectures for the Automated Diagnosis of Dyslexia via Visual Task Paradigm Analysis

M. Zaree [1], B. Arefi-Rad [2], M. Mohebbi [1,*], R. Rostami[3]

[1] Department of Biomedical Engineering, K. N. Toosi University of Technology, Tehran, Iran
[2] Department of Electrical Engineering, K. N. Toosi University of Technology, Tehran, Iran
[3] Department of Psychology, University of Tehran, Tehran, Iran

ARTICLE INFO

Article History:
Received 23 January 2025
Received in revised form 24 April 2025
Accepted 8 June 2025
Available online 16 June 2025

Keywords:
Dyslexia, Electroencephalography (EEG), Ensemble Learning, Machine Learning, Event-Related Potential (ERP).

ABSTRACT

Dyslexia is a neurobiological learning disability that fundamentally impairs a child's literacy acquisition, specifically manifesting as persistent deficits in reading and writing. Absent a timely diagnosis, the disorder precipitates profound psychological distress and academic marginalization for both the pediatric patients and their families. Furthermore, delayed intervention often results in cumulative achievement gaps that become increasingly difficult to bridge by secondary education. Consequently, early screening and clinical intervention are paramount to preserving student self-esteem and optimizing long-term academic trajectories. This study proposes an automated diagnostic framework utilizing a Bagging (Bootstrap Aggregating) ensemble learning approach to classify dyslexia in children. The methodology involves the rigorous preprocessing of electroencephalogram (EEG) signals recorded across a 19-channel montage. Feature extraction focused on the morphometry of Event-Related Potentials (ERPs), specifically quantifying the amplitude and latency of key components. To address the "curse of dimensionality" inherent in the high-dimensional feature set, Principal Component Analysis (PCA) was implemented for optimal feature reduction. To ensure the generalizability of the model and mitigate the risk of overfitting, a K-fold cross-validation strategy was employed during the training phase. Finally, the Bagging classifier was deployed to distinguish between dyslexic and neurotypical subjects. The proposed ensemble framework demonstrated robust performance, yielding an average classification accuracy of 90.6%. Notably, the model achieved a sensitivity rate of 100%, ensuring no dyslexic cases were omitted, and a specificity of 81.2%, reflecting its capability to accurately identify neurotypical controls.

1. INTRODUCTION

Dyslexia is a neurodevelopmental condition that disrupts reading, writing, and spelling abilities and is widely recognized as the most common learning disorder. According to the ICD-10 criteria [1], it is defined by persistent difficulties in processing written language despite normal visual function and adequate educational exposure. The disorder often exhibits a familial pattern, suggesting a hereditary component, and can persist into adulthood. In the

[*] Corresponding Author: m.mohebbi@kntu.ac.ir
Department of Biomedical Engineering, K. N. Toosi University of Technology, Tehran, Iran

http://dx.doi.org/10.22034/TMI.2025.244851

absence of timely intervention, dyslexia may result in academic underperformance and social withdrawal [2]. Importantly, dyslexia is not associated with intellectual deficits, as affected children generally possess cognitive abilities comparable to those of their peers. Epidemiological studies indicate that prevalence varies among English-speaking populations, affecting roughly 10% of individuals in Australia and up to 20% in Canada and the United Kingdom, with similar trends observed across other linguistic communities [3]. Beyond academic challenges, dyslexia can also contribute to negative emotional outcomes, including frustration, anxiety, depression, and diminished self-esteem.

Early and accurate identification is essential to support the educational and developmental needs of individuals with dyslexia. Conventional diagnostic procedures, however, are often resource-intensive, time-consuming, and rely heavily on reading and writing assessments, which may be unsuitable for pre-literate children. In recent years, technological advances have facilitated alternative diagnostic approaches, including machine learning frameworks, eye-tracking methodologies, and biomedical signal analysis. Among these, electroencephalography (EEG), which captures the brain’s electrical activity, has emerged as a non-invasive and objective tool for early detection. EEG-based assessment can reveal distinctive neural patterns associated with dyslexia, even prior to literacy acquisition.

Several studies have investigated EEG-driven techniques for dyslexia detection. For instance, Perera et al. employed EEG recordings combined with Support Vector Machines (SVM) to extract and classify relevant features, discussing the method’s strengths, limitations, and potential optimization strategies [4]. Kohli and Prasad proposed an Artificial Neural Network (ANN) model that effectively identified dyslexia and suggested underlying contributing factors [5]. Martins et al. developed a mobile application designed to diagnose dyslexia and support reading skill development, demonstrating robust usability and diagnostic performance [6].

Additional contributions include work by Shiratuddin et al., who analyzed pseudoword reading tasks using EEG power spectra-derived statistical features and classified results with an SVM approach [7]. Frid and Breznitz introduced an ensemble SVM algorithm capable of distinguishing dyslexic from typical readers based on EEG data, achieving high classification reliability, although their study focused solely on adults and did not account for related learning subtypes, such as dyscalculia or dysgraphia [8]. Furthermore, Cui et al. combined MRI with diffusion tensor imaging (DTI) to extract white matter characteristics, achieving an SVM classification accuracy of 83.61% in differentiating dyslexic children from control subjects [9].

Building upon this foundation, the present study aims to advance dyslexia detection through the analysis of Event-Related Potential (ERP) components during visual tasks. By extracting key ERP features, including amplitude and latency, and employing ensemble machine learning algorithms, this research seeks to accurately distinguish individuals with dyslexia from typically developing peers, contributing to more accessible, objective, and reliable diagnostic approaches.

2. METHOD

In this study, a machine learning framework was employed to detect dyslexia and investigate differences in EEG signals between individuals with and without the disorder. The methodology encompasses four primary stages: database description, signal pre-processing, feature extraction, and classification with performance optimization.

2.1. Dataset

The study population consisted of individuals referred to the Atiye Derakhshan Zehn Center due to challenges in reading, writing, and learning. The experimental sample comprised 44 participants, including 22 non-dyslexic individuals (10 females and 12 males) and 22 dyslexic individuals (4 females and 18 males). At the center, following clinical evaluation and diagnosis, participants’ brain activity was recorded. For individuals with dyslexia, EEG signals were collected under multiple conditions, including eyes-open, eyes-closed, and during engagement in visual tasks, using a 19-channel Mitsar electroencephalograph system. Figure 1 illustrates the electrode positions according to the international 10-20 system, along with their corresponding labels on the scalp [10].

2.2. Visual Task

EEG recordings primarily reflect postsynaptic membrane potentials, which are modulated by sensory, motor, and cognitive events. However, only a subset of these neuronal activities propagates through the skull and can be reliably measured. In this study, the Visual Continuous Performance Task (VCPT) was employed to elicit Event-Related Potentials (ERPs). A concise overview of this task, based on the methodology described by Brunner et al., is presented in the following section [11, 12].

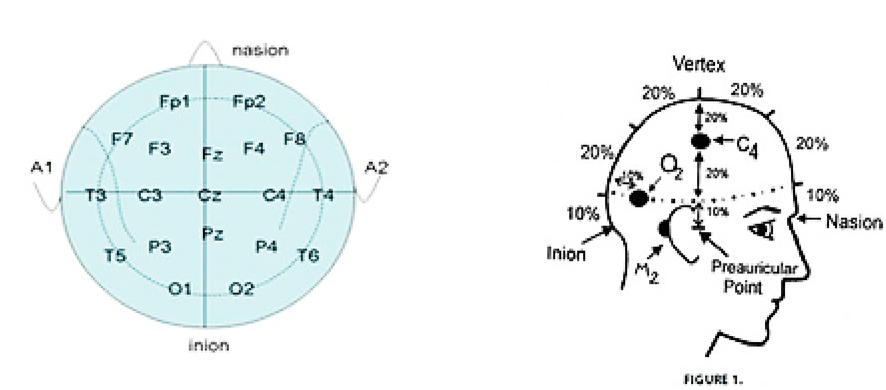

Fig. 1. Electrode Placement Based on the International 10-20 System [11].

The visual task employed in this study is a modified two-stimulus paradigm, incorporating three distinct categories of visual stimuli:

- 20 images depicting animals
- 20 images depicting plants
- 20 images depicting humans, accompanied by an artificial voice

The experimental procedure consists of 400 trials, presented in four different conditions with equal probability and in a randomized order, with a total duration of 20 minutes. Participants were provided with a 5-minute rest interval between each task block. Each trial involved the sequential presentation of two images, with each image displayed for 100 milliseconds and an inter-stimulus interval of 1000 milliseconds.

In this paradigm, trials beginning with an animal image required participants to be prepared for a response, whereas trials starting with a plant image required no response. Specifically, when the first image depicted an animal and the second image was also an animal, participants were instructed to press the mouse button with their right hand as quickly as possible. No response was required for any other combination of stimuli.

Figure 2 illustrates the four experimental conditions, labeled Event1 through Event4, corresponding to the different stimulus sequences from top to bottom.

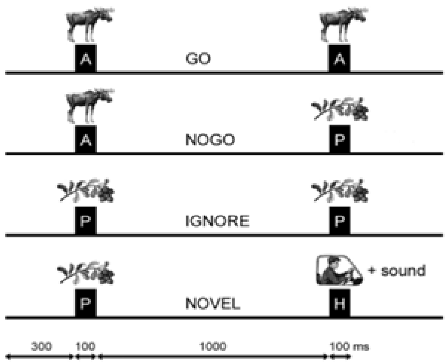

Fig. 2. Representation of the Four VCPT Events and Their Respective Stimulation Time Intervals [12].

2.3. Pre-Processing and Feature Extraction

2.3.1. Pre-Processing

EEG signals are inherently susceptible to noise and artifacts arising from environmental interference and physiological activities. Additional biological signals, including electrooculography (EOG) generated by eye movements and electromyography (EMG) from muscle contractions and respiration, are also recorded. Effective analysis of brain activity requires the removal of such confounding influences. Initially, a finite impulse response (FIR) filter was applied within the 0.5–45 Hz frequency range to attenuate undesired components. Subsequently, an average reference was computed by averaging all channels and subtracting this mean from each individual channel. To further minimize residual artifacts, particularly those associated with eye blinks and movements, Independent Component Analysis (ICA) was employed. ICA decomposes the EEG signal into statistically independent components, each representing neural activity from distinct brain regions.

To extract Event-Related Potentials (ERPs), EEG segments corresponding to stimulus-triggered events, known as epochs, were isolated. These epochs were then averaged to compute the ERP signal. For this study, epochs encompassed the full duration of stimulation, ranging from stimulus onset at 0 milliseconds to 3000 milliseconds post-stimulation.

2.3.2. ERP Extraction

EEG recordings capture a vast array of brain processes, making it challenging to isolate responses specific to individual stimuli or events. To address this, multiple repetitions of each experimental condition were conducted, and results were averaged to reduce random brain activity, thereby emphasizing the ERP waveform. From the ERP signal, key components including N100, P100, P200, P300, N170, N400, and the Late Positive Component (LPC) were identified across all channels. The amplitude and latency of each component were subsequently measured. Given that EEG data were recorded from 19 channels, a total of 266 features were computed for each participant.

2.3.3. Feature Selection and Dimensionality Reduction

Feature selection is a critical step in machine learning, as excessive features can increase computational demands and reduce classification accuracy. Moreover, high-dimensional feature spaces with limited samples can lead to dispersion and degraded model performance. To address this, the Kolmogorov-Smirnov (KS) test, a non-parametric statistical method, was employed to compare feature distributions between dyslexic and non-dyslexic groups. Features exhibiting statistically significant differences were retained, resulting in a subset of 12 relevant features.

Despite this reduction, the remaining 12 features were still considered high-dimensional for effective classification. Therefore, Principal Component Analysis (PCA) was applied to further reduce dimensionality [13]. PCA transforms the original correlated features into a smaller set of uncorrelated principal components that capture the majority of variance in the data. For classification purposes, the six most informative principal components were selected, providing a compact yet representative feature set for distinguishing dyslexic from non-dyslexic participants.

2.4. Classification

Bootstrap Aggregating, or bagging, is an ensemble learning strategy designed to enhance the stability and predictive performance of models, particularly those prone to high variance such as decision trees [14, 15]. Bagging generates multiple subsets of the training dataset via random sampling with replacement (bootstrapping), upon which independent base classifiers are trained. For classification tasks, final predictions are obtained through majority voting, whereas regression outputs are averaged. This approach reduces overfitting by decreasing model variance and smoothing out noise, thereby improving generalization.

In this study, bagging was applied to several classifiers, including decision trees, non-linear Support Vector Machines (SVM), Naive Bayes, and K-Nearest Neighbors (KNN). Among these, the non-linear SVM demonstrated the highest accuracy and sensitivity. Ensemble predictions from the bagging framework were combined via majority voting, resulting in significant improvements in classification performance, robustness, and resistance to overfitting. The subsequent sections present a detailed evaluation of the method's effectiveness.

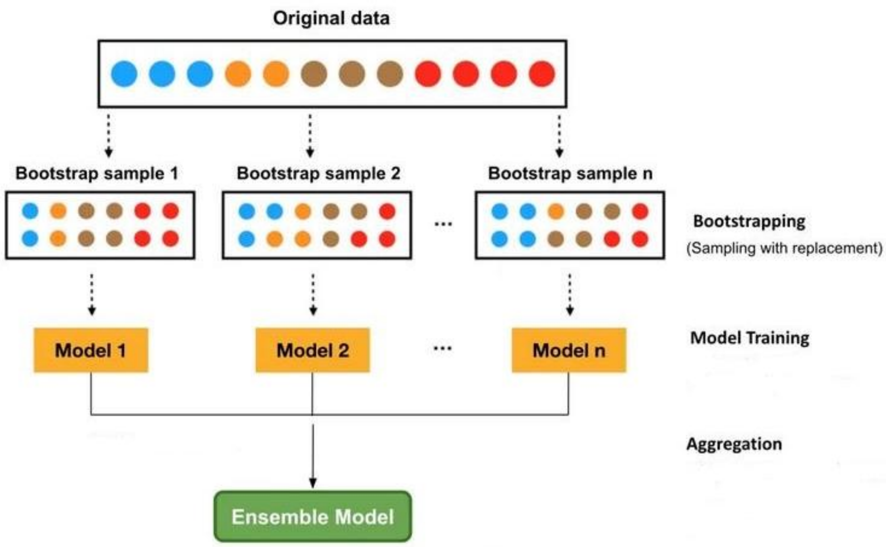

Fig. 3. Bagging methods of ensemble learning [16]

3. RESULTS

Figure 4 presents a block diagram illustrating the overall process employed to differentiate between individuals with dyslexia and those without.

Fig. 4. Overview of the algorithm's block diagram presented in this research.

After recording and preprocessing the EEG signals, ERP components were extracted from each channel, and their amplitudes and latencies were calculated as signal features. This process yielded a total of 266 features per person. Using Ks-test, we initially selected the top 12 features for classification. However, the classification results were not satisfactory with this feature set. Consequently, we applied PCA for dimensionality reduction and selected 6 components from the reduced feature set. In the final step, we employed the bagging method within ensemble learning to classify the data. Table 1 presents the results of classification and cross-validation with K=8 for all four modes of the VCPT assignment. As shown in the table, for Event 2, we achieved an accuracy of 90.6%, sensitivity of 100%, and specificity of 81.2% which are strong and acceptable results for diagnosing this disorder.

Table 1. Classification Results for 4 Events in VCPT Task.

	Accuracy%	Sensitivity%	Specificity%	F1-Score
Event 1	68.7	62.5	75	83.2
Event 2	**90.6**	**100**	**81.2**	**91.2**
Event 3	62.5	56.2	68.7	72.1
Event 4	66.6	77.7	55.5	65.8

4. DISCUSSION

In the present study, following EEG preprocessing and ERP computation through averaging, key components including P100, N100, N170, P200, P300, N400, and the Late Positive Component (LPC) were extracted from all channels. The amplitude and latency of each component were quantified, after which the five most discriminative features were selected using the Kolmogorov-Smirnov (KS) test and Principal Component Analysis (PCA). Classification was then performed using a bagging ensemble approach. Analysis of the selected features revealed that the majority of relevant electrodes were located in the left hemisphere, which encompasses critical language-related regions such as Wernicke's and Broca's areas. This lateralization suggests that dyslexic individuals may exhibit deficits in language processing and writing skills.

Classification results indicated that the proposed algorithm achieved the highest performance during Event 2. In this condition, participants were required to process the initial stimulus and prepare to respond appropriately by pressing the mouse button if the stimulus depicted an animal, while inhibiting responses when it did not. This task specifically assesses the ability to suppress prepotent responses. The observed performance deficits in dyslexic participants suggest challenges in response inhibition, reflecting difficulties in controlling initial impulses and in accurately and promptly analyzing incoming stimuli. Such impairments in impulse control and stimulus processing appear to be characteristic of dyslexia in this experimental context.

Additionally, classification accuracy was also higher in Event 1 relative to other conditions. Event 1 involved the initial stage of stimulus recognition, requiring participants to respond based on rapid identification of the presented image. The reduced performance of dyslexic individuals in this event highlights deficits in early cognitive processing stages, where swift and accurate recognition is essential. These findings collectively underscore significant differences between dyslexic and non-dyslexic participants in attentional focus, recognition speed, and analytical processing. Dyslexic individuals typically demonstrate slower cognitive processing and challenges in maintaining sustained attention, which adversely affect performance in tasks demanding rapid and precise responses.

Considering that pre-school children generally have underdeveloped reading and writing skills, ERP components elicited during the Visual Continuous Performance Task (VCPT) provide valuable neural markers for early dyslexia detection. These components reflect underlying cognitive processes, including attention allocation, perceptual processing, and decision-making. By analyzing the amplitude and latency patterns of these ERP components, it is possible to identify neural signatures that reliably differentiate dyslexic children from their peers. Early identification enables the design of targeted interventions aimed at specific cognitive and neural deficits, thereby improving educational outcomes and overall quality of life. Timely, individualized intervention informed by ERP analysis can enhance the effectiveness of dyslexia management and support the development of essential literacy skills.

Declaration

We acknowledge that we used ChatGPT to enhance the academic writing of our manuscript while ensuring the originality and integrity of our work.

Transparency Statement

The data supporting this study are available upon reasonable request to the corresponding author, subject to ethical and confidentiality considerations.

Acknowledgments

We would like to express our gratitude to all individuals who contributed to this project.

Declaration of Interest

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

REFERENCES

[1] Gagliardi, L., Rusconi, F., Bellù, R., Zanini, R., & Italian Neonatal Network. (2014). Association of maternal hypertension and chorioamnionitis with preterm outcomes. *Pediatrics*, 134(1), e154–e161. https://doi.org/10.1542/peds.2013-3898.

[2] Zahia, S., Garcia-Zapirain, B., Saralegui, I., & Fernandez-Ruanova, B. (2020). Dyslexia detection using 3D convolutional neural networks and functional magnetic resonance imaging. *Computer Methods and Programs in Biomedicine*, 197, 105726. https://doi.org/10.1016/j.cmpb.2020.105726

[3] Rauschenberger, M., Rello, L., Baeza-Yates, R., & Bigham, J. P. (2018, April). Towards language independent detection of dyslexia with a web-based game. In *Proceedings of the 15th International Web for All Conference* (pp. 1–10). https://doi.org/10.1145/3192714.3192816

[4] Perera, H., Shiratuddin, M. F., & Wong, K. W. (2016). A review of electroencephalogram-based analysis and classification frameworks for dyslexia. In *Neural Information Processing: 23rd International Conference, ICONIP 2016, Proceedings, Part IV* (pp. 626–635). Springer. https://doi.org/10.1007/978-3-319-46681-1_74

[5] Kohli, M., & Prasad, T. V. (2010, June). Identifying dyslexic students by using artificial neural networks. In *Proceedings of the World Congress on Engineering* (Vol. 1, pp. 1–4). WCE.

[6] Martins, V. F., Lima, T., Sampaio, P. N., & de Paiva, M. (2016, November). Mobile application to support dyslexia diagnostic and reading practice. In *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)* (pp. 1–6). IEEE. https://doi.org/10.1109/AICCSA.2016.7945710

[7] Perera, H., Shiratuddin, M. F., Wong, K. W., & Fullarton, K. (2017, June). EEG signal analysis of real-word reading and nonsense-word reading between adults with dyslexia and without dyslexia. In *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 73–78). IEEE. https://doi.org/10.1109/CBMS.2017.108

[8] Frid, A., & Breznitz, Z. (2012, November). An SVM based algorithm for analysis and discrimination of dyslexic readers from regular readers using ERPs. In *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel* (pp. 1–4). IEEE. https://doi.org/10.1109/EEEI.2012.6377068

[9] Cui, Z., Xia, Z., Su, M., Shu, H., & Gong, G. (2016). Disrupted white matter connectivity underlying developmental dyslexia: A machine learning approach. *Human Brain Mapping*, 37(4), 1443–1458. https://doi.org/10.1002/hbm.23112

[10] Meissner, N. A. (2018). *A single-subject evaluation of facilitated communication in the completion of school-assigned homework* [Master's thesis, University of Wisconsin-Eau Claire]. Minds@UW.

[11] Heshmatollah, A., Dommershuijsen, L. J., Fani, L., Koudstaal, P. J., Ikram, M. A., & Ikram, M. K. (2021). Long-term trajectories of decline in cognition and daily functioning before and after stroke. *Journal of Neurology, Neurosurgery & Psychiatry*, 92(11), 1158–1163. https://doi.org/10.1136/jnnp-2021-326043

[12] Kropotov, J. D., & Ponomarev, V. A. (2009). Decomposing N2 NOGO wave of event-related potentials into independent components. *Neuroreport*, 20(18), 1592–1596. https://doi.org/10.1097/WNR.0b013e3283309cbd

[13] Daffertshofer, A., Lamoth, C. J., Meijer, O. G., & Beek, P. J. (2004). PCA in studying coordination and variability: A tutorial. *Clinical Biomechanics*, 19(4), 415–428. https://doi.org/10.1016/j.clinbiomech.2004.01.005

[14] Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. https://doi.org/10.1007/BF00058655

[15] Altman, N., & Krzywinski, M. (2017). Ensemble methods: Bagging and random forests. *Nature Methods*, 14(10), 933–935. https://doi.org/10.1038/nmeth.4438

[16] AIML. (2022). *What is bagging?* https://aiml.com/what-is-bagging/.