

Article

Not peer-reviewed version

Developing a Dyslexia Indicator Using Eye Tracking

Kevin Cogan , [Vuong M. Ngo](#) ^{*} , [Mark Roantree](#)

Posted Date: 6 May 2025

doi: 10.20944/preprints202505.0310.v1

Keywords: eye-tracking; dyslexia detection; reading difficulties; early diagnosis; Machine Learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Developing a Dyslexia Indicator Using Eye Tracking

Kevin Cogan ^{1,†}, Vuong M. Ngo ^{2,†,*}  and Mark Roantree ³ 

¹ School of Computing, Dublin City University, Ireland

² Ho Chi Minh City Open University, Ho Chi Minh City, Vietnam

³ Insight Centre for Data Analytics, Dublin City University, Ireland

* Correspondence: vuong.cs@gmail.com

† These authors designated as co-first authors.

Abstract: Dyslexia, affecting an estimated 10% to 20% of the global population, significantly impairs learning capabilities, highlighting the need for innovative and accessible diagnostic methods. This paper investigates the effectiveness of eye-tracking technology combined with machine learning algorithms as a cost-effective alternative for early dyslexia detection. By analyzing general eye movement patterns, including prolonged fixation durations and erratic saccades, we proposed an enhanced solution for determining eye-tracking-based dyslexia features. A Random Forest Classifier was then employed to detect dyslexia, achieving an accuracy of 88.58%. Additionally, hierarchical clustering methods were applied to identify varying severity levels of dyslexia. The analysis incorporates diverse methodologies across various populations and settings, demonstrating the potential of this technology to identify individuals with dyslexia, including those with borderline traits, through non-invasive means. Integrating eye-tracking with machine learning represents a significant advancement in the diagnostic process, offering a highly accurate and accessible method in clinical research.

Keywords: eye-tracking; dyslexia detection; reading difficulties; early diagnosis; Machine Learning

1. Introduction

Dyslexia is commonly understood as a specific learning disability, characterized by difficulties with accurate and/or fluent word recognition, alongside poor spelling and decoding abilities which result in challenges with reading fluency, comprehension, and vocabulary growth [1]. Neurologically, dyslexia is associated with differences in the brain's structure and function. Studies using functional magnetic resonance imaging (fMRI) and other neuroimaging techniques have identified atypical activations in regions of the brain responsible for key reading processes, including the left temporal-parietal cortex, which is involved in phonological processing and decoding [2].

Dyslexia poses significant challenges in educational environments, impacting individuals' reading, writing, and language processing abilities. It is estimated to affect 10% to 20% of the global population [3]. Early detection is crucial for providing effective interventions, yet many face difficulties due to the high costs, limited access to specialized psychological assessments, and the heavy reliance on teachers to quickly identify learning difficulties.

Eye-tracking technology offers a non-invasive method to understand dyslexia by monitoring rapid eye movements, such as saccades and fixations. In the context of dyslexia, eye movement patterns differ significantly from those of non-dyslexic readers [4]. Individuals with dyslexia often exhibit longer fixation durations, more frequent and shorter saccades, and a higher number of regressions, reflecting difficulties in processing and decoding text. These distinctive eye movement patterns serve as biomarkers for dyslexia, allowing researchers and clinicians to explore the neurocognitive discrepancies that characterize the disorder.

Dyslexia diagnosis has evolved from subjective methods to advanced technologies. The 1990s brought costly neuroimaging tools like fMRI and PET, revealing neurological differences [5]. By the 2010s, eye-tracking and machine learning emerged as efficient, non-invasive tools for analyzing

reading behavior and identifying dyslexic traits [6]. Machine learning algorithms now enable the automatic classification of dyslexic and non-dyslexic readers based on eye movement data during reading [7–10]. This innovative approach leverages computational power to provide early, accurate, and efficient dyslexia diagnosis, facilitating timely interventions and support.

Contribution. This paper investigates the use of eye-tracking technology as an innovative and cost-effective solution for dyslexia detection. By capturing precise eye movement metrics, such as saccades and fixations, eye-tracking offers a non-invasive and accessible method for identifying individuals with dyslexia. Building on foundational eye-tracking features from existing research, we propose an enhanced framework for developing eye-tracking-based dyslexia features. Machine learning techniques, including Random Forest Classifier and Agglomerative Hierarchical Clustering, are applied to analyze dyslexia-related patterns. The study evaluates the accuracy, effectiveness, and potential of eye-tracking in early dyslexia detection, highlighting its ability to complement traditional diagnostic methods. Additionally, it reviews research demonstrating the scalability and adaptability of eye-tracking across diverse populations and educational contexts, emphasizing its suitability for large-scale screening initiatives. By integrating advanced technologies, this paper aims to promote more inclusive, timely, and efficient dyslexia assessments.

Paper Structure. The remainder of this paper is organized as follows: section 2 discusses related work; section 3 outlines our data modeling approach and describes the solution for extracting enhanced eye-tracking-based dyslexia features, which are then used to apply machine learning algorithms; the experimental results and analyses for both classification and clustering tasks are presented in section 4; and finally, section 5 concludes the paper and suggests directions for future work.

2. Related Work

The work presented in [11] focuses on the comparison of reading strategies between individuals with dyslexia and non-dyslexic controls using eye movement analysis during standardized reading tasks. This study utilized the International Reading Speed Texts (IReST) for consistency in material complexity, with eye-tracking technology capturing critical metrics such as fixation duration, saccade length, and regression frequency. This approach sheds light on the visual sampling strategies employed by individuals with dyslexia.

Building on the application of eye-tracking, [12] integrates machine learning techniques to diagnose developmental dyslexia, leveraging eye movement data from 165 participants in the eSeek project. Dyslexic individuals were identified based on low reading fluency scores, and the methodology employed Random Forest for feature selection and Support Vector Machines for classification. This dual approach effectively reduced the data's dimensionality while addressing its complexity, underscoring the potential of machine learning for handling high-dimensional datasets in dyslexia research.

Similarly, [13] introduces the VGG16 model to analyze eye-tracking data for dyslexia detection in children aged 7 to 13. Unlike traditional preprocessing methods, this study transformed gaze data into color-coded 2D time series graphs, capturing intricate eye movement patterns. These graphs served as input for classification, with the study further exploring various network configurations and segmentation strategies to enhance the model's accuracy.

Building on these advancements, Nerusil et al. [6] propose a groundbreaking method for dyslexia detection, utilizing Convolutional Neural Networks (CNNs) to process eye-tracking data holistically. The dataset comprises eye-tracking signals from 185 subjects, classified into low and high-risk categories for dyslexia. Departing from traditional methods, this approach minimizes preprocessing and eliminates explicit feature extraction, allowing the CNN to autonomously detect and classify dyslexia-related patterns directly from raw data. The signals are analyzed in both time and frequency domains, with data standardized through zero-padding or interpolation for time-domain representation and magnitude spectrum analysis for frequency-domain analysis.

Finally, [4] takes a feature-based machine learning approach, combining conventional eye-tracking metrics with novel features to characterize reading behaviors. The inclusion of innovative measures,

such as the Fixation Intersection Coefficient (FIC) and Fixation Fractal Dimension (FFD), offers deeper insights into the spatial and temporal gaze dynamics associated with dyslexia. By employing various machine learning algorithms for classification, this study enriches our understanding of dyslexic reading patterns and their distinct characteristics.

In summary, the studies mentioned above did not incorporate a sufficient number of eye-tracking-based dyslexia features into their machine learning models, which likely limited the classification performance. Additionally, none of these projects focused on analyzing the correlations between features and clusters, missing an opportunity to extract more meaningful insights. Without this correlation analysis, valuable relationships between different eye-tracking metrics and dyslexia severity could not be fully leveraged to enhance the accuracy and interpretability of the models. We believe that by incorporating a broader set of features and performing correlation analyses, this offers the opportunity to significantly improve model performance and provide deeper insights into the underlying patterns of dyslexia.

3. Feature Engineering

The Provo Corpus comprises quantitative eye-tracking data collected from participants at Brigham Young University [14]. The dataset, approximately 78MB in size, contains 230,412 entries spanning 63 columns of both numerical and categorical data. It includes detailed records of eye movements from proficient readers interacting with continuous text, along with predictability scores derived from the Predictability Norms surveys. These scores incorporate traditional cloze measures, as well as morpho-syntactic and semantic predictability metrics for each word. This dataset provides detailed insights into reading behaviors, such as fixations, saccades, and regressions. During the eye-tracking sessions, participants' eye movements were recorded using an SR Research EyeLink 1000 Plus eye-tracker, enabling precise measurement of their reading dynamics.

3.1. Dataset Description

Predictability Norms Data. Predictability norms data were collected through a large-scale online survey involving 470 participants who were randomly assigned to complete five out of 55 texts. This helped quantify the predictability of each word within its context. The texts averaged 50 words and approximately 2.5 sentences each. In total, the dataset included 2,689 words, with 1,197 being unique word forms. Participants also completed an online survey via Qualtrics, where they provided demographic information and responded to text completion tasks. The texts were tagged for parts of speech using the Constituent Likelihood Automatic Word-Tagging System (CLAWS), and semantic association scores for each word were derived through latent semantic analysis (LSA).

Eye-Tracking Data. Eye-tracking data were gathered from 84 native English-speaking participants who read the same 55 texts used in the Predictability Norms group. The texts were sourced from a range of materials, including online news articles, popular science magazines, and public-domain fiction.

3.2. Data Processing

Data cleaning: To streamline the model and improve its performance, redundant columns with high missing values, such as `Ia_Regression_Path_Duration` and `Ia_First_Run_Dwell_Time`, were removed, along with highly correlated features like `Ia_First_Saccade_Start_Time` and `Ia_First_Saccade_End_Time`. Then, missing values were addressed using advanced techniques: the MissForest algorithm¹ was used to impute missing `Ia_Skip` values, rows lacking the `Word` value were excluded, and sequential fields like `Sentence_Number` and `Word_In_Sentence_Number` were filled using the forward fill method. Integrating semi-structured data with categorical and continuous variables used the method presented in [15]. Finally, to manage outliers and ensure consistency, selected

¹ <https://pypi.org/project/MissForest/>

columns were normalized using MinMax Scaling, which adjusts the data to a specified range (0 to 1), enhancing the model's accuracy and robustness.

Feature transformation: We developed a new feature, *Saccade_Duration*, by calculating the time difference between *Ia_First_Saccade_End_Time* and *Ia_First_Saccade_Start_Time*, to more accurately measure the duration between saccades. Additionally, we transformed the *Word_Cleaned* column using the TF-IDF technique. This transformation enhances the analytical significance of each word by quantifying its importance relative to the document corpus, enabling more nuanced text analysis.

3.3. Enhanced Eye-Tracking-based Dyslexia Feature Determination

a) Basic Dyslexia Features:

Our research focuses on uncovering distinctions in eye-tracking metrics between individuals with and without dyslexia, building on insights from existing studies. By analyzing these differences through both statistical methods and visualizations—particularly by examining how these metrics vary across percentiles—we aim to identify meaningful patterns that can effectively distinguish skilled readers from those facing reading challenges. This dual approach not only highlights significant statistical disparities but also provides an intuitive understanding of the variations in reading behavior. By identifying how these metrics differ across readers, our study provides valuable insights into the cognitive and behavioral characteristics of dyslexia. Key metrics considered include:

- *Ia_First_Saccade_Amplitude*: This metric quantifies the distance of the initial eye movement. Typically, dyslexic individuals exhibit shorter amplitudes, suggesting less efficient initial scanning of text [11].
- *Ia_Dwell_Time*: This records the total time eyes fixate on a word. Extended dwell times in dyslexic readers imply greater effort in word recognition and processing [11].
- *Ia_Regression_In_Count* and *Ia_Regression_Out_Count*: These metrics count the instances where a reader's gaze regresses to previously read text, with dyslexics displaying more frequent regressions, a marker of disrupted reading fluency [16].
- *Ia_Fixation_Count*: Captures the frequency of eye fixations on text, where higher counts are indicative of increased cognitive effort during reading, a common trait among dyslexics [17].
- *Saccade_Duration*: Measures the duration of eye movements between fixations. Shorter durations in dyslexic readers reflect quick but inefficient scanning [11].

Thresholds for labeling: The 95th percentile was used as the threshold for each metric to differentiate between levels of reading difficulty. Any data entry above this threshold was assigned a label of 1, indicating a higher level of reading difficulty, while entries below were labeled 0, indicating lower difficulty. These labels were recorded in a new column named *Reading_Difficulties* [12].

b) Enhancing Dyslexia Features:

To determine the key factors influencing the *Reading_Difficulty* feature, a *RandomForestClassifier* was applied to a balanced dataset. The dataset was balanced by resampling the majority class to correct class imbalances, ensuring fairer model training and evaluation. Features directly contributing to the construction of the *Reading_Difficulty* target, such as *Ia_First_Saccade_Amplitude*, *Ia_Dwell_Time*, and *Ia_Fixation_Count*, were excluded from the analysis to avoid data leakage and ensure an unbiased assessment of feature importance.

The model was optimized using Bayesian hyperparameter tuning, which systematically explored the parameter space to maximize the classifier's performance. Once trained, the model identified the most influential features contributing to the prediction of reading difficulty. These key features were visualized using a bar plot, providing a clear, data-driven understanding of their relative importance and offering insights for refining the model further.

The most impactful features for predicting reading difficulty included: *Ia_Regression_In_Count*, *Ia_First_Run_Fixation_.*, *Saccade_Duration*,

Ia_Right, Ia_First_Fixation_Time, Ia_Regression_Out_Count, Ia_First_Fixation_Index, Ia_First_Fixation_Y, Ia_First_Fixation_X, Word_Number, Ia_Skip and Word_Length. These features, alongside the basic eye-tracking metrics described in Section 4a, form a comprehensive set of eye-tracking-enhanced dyslexia features. This enriched feature set aims to improve the detection and understanding of reading difficulties, particularly in dyslexic individuals.

3.4. Clustering and Classification Algorithms

a) Random Forest Classifier:

Random Forest classification creates an ensemble of decision trees, each trained on a random subset of data through bootstrap sampling. At each node, a random subset of features is evaluated for splitting, promoting diversity and preventing any single feature from dominating the model. Each tree grows independently and makes its own prediction. For classification, the final prediction is based on majority voting across all trees [18].

The hyperparameters of the Random Forest classifier is fine-tuned using BayesSearchCV, concentrating on the features selected during the forward selection process. This approach aims to optimize the cross-validation score by exploring various hyperparameter configurations. The Random Forest model is refined with key features identified through forward selection and benchmarked against a model trained on all features, with both employing optimized hyperparameters. To ensure robustness and manage variability, the model was trained on multiple resampled data subsets to address class imbalance.

b) Agglomerative Hierarchical Clustering:

Agglomerative Hierarchical Clustering (AHC) is a bottom-up clustering method where each data point begins as its own cluster, and pairs of clusters are merged iteratively based on similarity until all data points belong to a single cluster or a predefined number of clusters are reached. We employ Euclidean distance and Ward Linkage to calculate distances between clusters and determine how clusters are combined [19].

One advantage of AHC is that it does not require the number of clusters to be specified beforehand, offering flexibility in exploratory data analysis. So, AHC is widely used in fields like bioinformatics, marketing, and text analysis due to its interpretability and ability to capture nested groupings of data points. Principal Component Analysis (PCA) [20] is applied to reduce the dataset to two principal components. This reduction was essential for visualizing the clusters in a two-dimensional plot, providing immediate and clear visual insights into the complex patterns within the data.

4. Evaluation and Discussion

In this section, we present a two-step validation process. First, we evaluate the performance of our dyslexia classifier, which identifies whether a reader likely displays dyslexia-related patterns. Second, we use cluster analysis to group readers by shared eye-tracking characteristics, offering deeper insight into the variability and potential severity of reading difficulties. Together, these complementary methods, classification for detection and clustering for subgroup analysis, deliver both strong diagnostic accuracy and a richer understanding of reading behavior, thereby reinforcing the robustness and interpretability of our overall findings.

4.1. Classification Result

The evaluation of the Random Forest Classifier was conducted using a 9-fold cross-validation process to ensure robustness and generalizability across various data subsets. This comprehensive evaluation aimed to establish a reliable baseline for the model's effectiveness in operational settings, with a focus on consistent performance metrics. Table 1 presents the average metrics from the 9-fold cross-validation, reporting an accuracy of 88.58%, precision of 87.91%, recall of 89.49%, and an F1-Score of 88.67%.

Table 1. Average 9-Fold Cross-Validation Metrics

Metric	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Average	88.58	87.91	89.49	88.67

Additionally, Figure 1(a) displays the average confusion matrix, further highlighting the Random Forest Classifier’s consistent performance across all validation folds. The model demonstrates a strong balance between sensitivity (high true positive rate) and specificity (high true negative rate), confirming its reliability and effectiveness in practical applications. This balance is particularly critical in scenarios where accurate decision-making is essential, and the cost of errors is high, making the model suitable for deployment in such environments.

To further evaluate the model’s performance, the Receiver Operating Characteristic (ROC) curve was plotted for each cross-validation fold, and the average ROC curve is presented below. This curve illustrates the trade-offs between true positive rates (sensitivity) and false positive rates for different threshold settings. Figure 1(b) displays the average Area Under the Curve (AUC) for the ROC curve, which is 0.96, indicating an excellent level of discrimination between the positive and negative classes. The consistency of the ROC curves across the nine folds suggests that the model’s performance is robust and reliable across different data subsets. This high and consistent AUC value highlights the model’s suitability for practical applications, such as diagnosing reading difficulties, where both high sensitivity and specificity are essential.

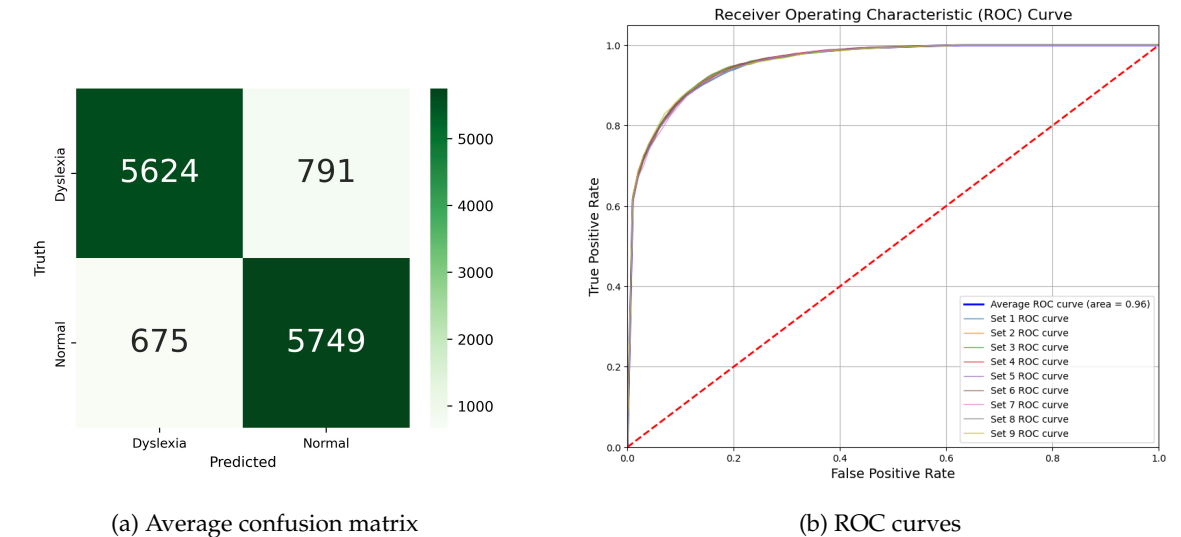


Figure 1. Experimental Results on 9-Fold Cross-Validation.

4.2. Cluster and Analysis

Figure 2 presents a visualization of clustering results based on the enhanced eye-tracking-based dyslexia features, revealing three distinct clusters. Notably, there is a significant overlap between Clusters 0 and 1, indicating some shared characteristics between the groups. The detailed profiling of each cluster is as follows:

- Cluster 0: This cluster is characterized by low mean and median dwell times, shorter saccade durations, and moderate regression counts. These patterns suggest readers with efficient eye movement strategies, indicative of proficient readers who process text with minimal cognitive effort and high fluency.
- Cluster 1: Readers in this cluster exhibit higher dwell times, lower regression counts, and average saccade durations. These metrics reflect moderately proficient or average readers who may occasionally encounter difficulties but generally exhibit effective reading behaviors. This group likely includes readers with slightly reduced fluency compared to Cluster 0 but without significant impairments.

- Cluster 2: This cluster is marked by significantly higher saccade amplitudes, increased dwell times, elevated fixation counts, and longer saccade durations. These patterns suggest a group of poor readers, potentially including individuals with dyslexia, who exhibit substantial effort and difficulty while reading. Their eye-tracking metrics indicate challenges in processing text efficiently, requiring more time and cognitive resources to decode words.

These findings provide a clear spectrum of reading behaviors, ranging from proficient to poor or dyslexic readers. The clustering results highlight the nuanced differences in eye movement patterns across the groups, with Cluster 2 particularly standing out for its association with reading difficulties. This visualization and clustering analysis underscore the potential of enhanced eye-tracking features to differentiate reading abilities. Such insights could inform the development of diagnostic tools for identifying dyslexia and tailoring targeted interventions to support individuals across the reading spectrum.

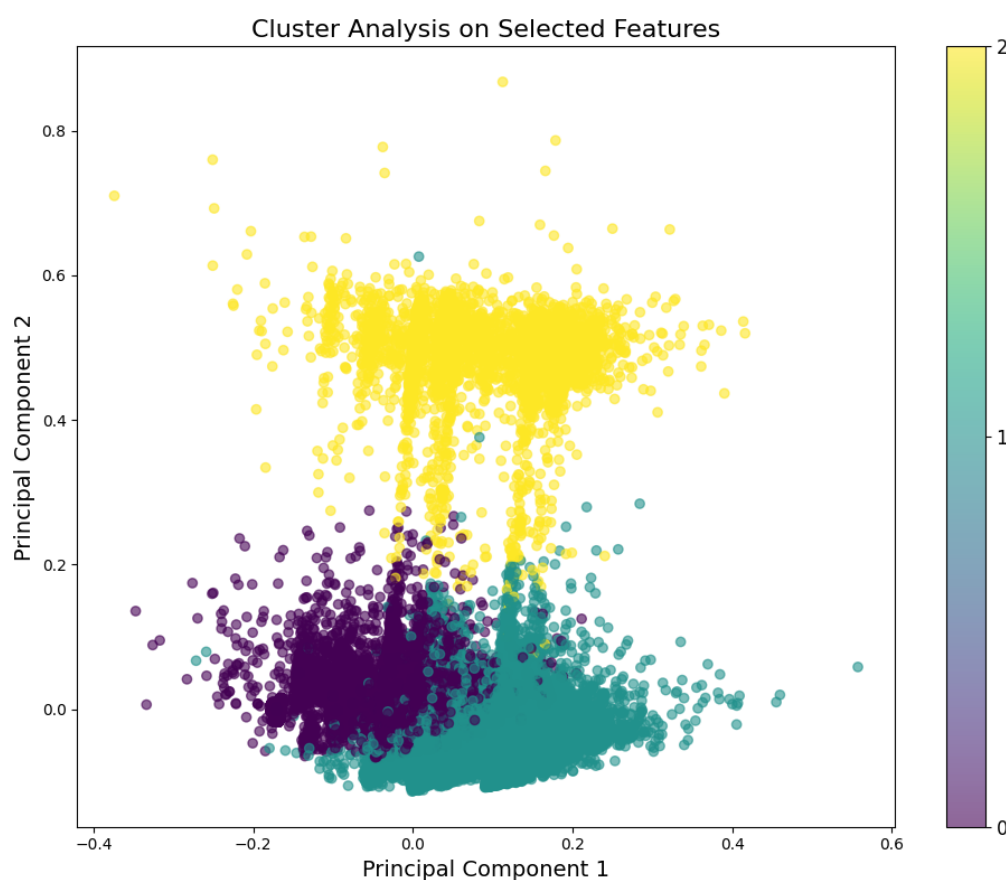


Figure 2. Cluster Analysis on Our Enhancing Dyslexia Features.

5. Conclusion and Future Work

The paper highlights the use of eye-tracking and machine learning to effectively detect dyslexia, leveraging metrics like saccade amplitude, dwell time, and fixation count. Operating at the 95th percentile, the method offers a non-invasive, objective, and cost-effective diagnostic tool, reducing reliance on subjective teacher judgment. This approach supports early identification of students needing intervention, enabling timely and targeted support [21].

The development and evaluation of the Random Forest classifier underscored the pivotal roles of hyperparameter tuning and feature selection. Hyperparameter tuning optimized the model's training process, enabling it to achieve superior results more efficiently. Selecting the top 12 essential features further enhanced performance while offering the potential to simplify future data collection and processing. The clustering analysis revealed distinct reading behaviors and strategies, illustrating a spectrum ranging from proficient readers to struggling ones, including individuals with dyslexic

traits. These findings highlight the diversity in reading approaches and underscore the need for further research into targeted interventions and specialized support by identifying shared characteristics within each group.

Future research could focus on integrating eye-tracking data with other diagnostic methods, such as neuroimaging and behavioral assessments, to develop more comprehensive and multimodal approaches to dyslexia diagnosis. Combining these modalities would provide a richer understanding of the neural, cognitive, and behavioral mechanisms underlying dyslexia. For instance, neuroimaging could reveal brain activity patterns associated with reading difficulties, while behavioral assessments could capture additional contextual factors, such as phonological processing and working memory. Additionally, a broader range of approaches—including Ontology and Knowledge Graph techniques [22], Natural Language Processing methods [23], and Machine Learning/Deep Learning models [24]—should be further applied and evaluated in this context.

Acknowledgments: This research has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland under Grant number 12/RC/2289_P2.

References

1. Reid, G. *Dyslexia: A Practitioner's Handbook, 5th Edition*; Wiley-Blackwell, 2016.
2. Bach, S.; Brandeis, D.; Hofstetter, C.; Martin, E.; Richardson, U.; Brem, S. Early Emergence of Deviant Frontal fMRI Activity for Phonological Processes in Poor Beginning Readers. *NeuroImage* **2010**, *53*, 682–693. <https://doi.org/10.1016/j.neuroimage.2010.06.039>.
3. Wu, Y.; Cheng, Y.; Yang, X.; Yu, W.; Wan, Y. Dyslexia: A Bibliometric and Visualization Analysis. *Frontiers in Public Health* **2022**, *10*. <https://doi.org/10.3389/fpubh.2022.915053>.
4. Smyrnakis, I.; Andreadakis, V.; Rina, A.; Boufachtrentin, N.; Aslanides, I. Silent versus Reading Out Loud modes: An eye-tracking study. *Journal of Eye Movement Research* **2021**, *14*, 10.16910/jemr.14.2.1. <https://doi.org/10.16910/jemr.14.2.1>.
5. Richards, T.; Berninger, V. Abnormal fMRI Connectivity in Children with Dyslexia During a Phoneme Task: Before but Not After Treatment. *Journal of Neurolinguistics* **2008**, *21*, 294–304. <https://doi.org/10.1016/j.jneuroling.2007.07.002>.
6. Nerušil, B.; Polec, J.; Škunda, J.; Kačur, J. Eye Tracking Based Dyslexia Detection Using a Holistic Approach. *Scientific Reports* **2021**, *11*. <https://doi.org/10.1038/s41598-021-95275-1>.
7. Poornappriya, T.; Gopinath, R. Application of Machine Learning Techniques for Improving Learning Disabilities, 2020. <https://doi.org/10.17605/OSF.IO/39ZQD>.
8. Vajs, I.; Ković, V.; Papić, T.; Savić, A.; Janković, M. Spatiotemporal Eye-Tracking Feature Set for Improved Recognition of Dyslexic Reading Patterns in Children. *Sensors* **2022**, *22*, 4900–4900. <https://doi.org/10.3390/s22134900>.
9. Vajs, I.; Savić, A.; Ković, V.; Janković, M.; Papić, T. Dyslexia Detection in Children Using Eye Tracking Data Based on VGG16 Network, n.d.
10. Vehlen, A.; Spenthof, I.; Tönsing, D.; Heinrichs, M.; Domes, G. Evaluation of an Eye Tracking Setup for Studying Visual Attention in Face-to-Face Conversations. *Scientific Reports* **2021**, *11*. <https://doi.org/10.1038/s41598-021-81987-x>.
11. Franzen, L.; Stark, Z.; Johnson, A. Individuals with Dyslexia Use a Different Visual Sampling Strategy to Read Text. *Scientific Reports* **2021**, *11*. <https://doi.org/10.1038/s41598-021-84945-9>.
12. Raatikainen, P.; Hautala, J.; Loberg, O.; Kärkkäinen, T.; Leppänen, P.; Nieminen, P. Detection of Developmental Dyslexia with Machine Learning Using Eye Movement Data. *Array* **2021**, *12*, 100087–100087. <https://doi.org/10.1016/j.array.2021.100087>.
13. Vajs, I.; Ković, V.; Papić, T.; Savić, A.; Janković, M. Dyslexia Detection in Children Using Eye Tracking Data Based on VGG16 Network. In Proceedings of the 2022 30th European Signal Processing Conference (EUSIPCO), 2022. <https://doi.org/10.23919/eusipco55093.2022.9909817>.
14. Luke, S.; Christianson, K. The Provo Corpus: A Large Eye-Tracking Corpus with Predictability Norms. *Behavior Research Methods* **2017**, *50*, 826–833. <https://doi.org/10.3758/s13428-017-0908-4>.
15. Scriney, M.; McCarthy, S.; McCarren, A.; Cappellari, P.; Roantree, M. Automating Data Mart Construction from Semi-structured Data Sources. *The Computer Journal* **2018**, *62*, 394–413,

- [<https://academic.oup.com/jnl/article-pdf/62/3/394/28009314/bxy064.pdf>]. <https://doi.org/10.1093/comjnl/bxy064>.
16. Vagge, A.; Cavanna, M.; Traverso, C.; et al. Evaluation of Ocular Movements in Patients with Dyslexia. *Annals of Dyslexia* **2015**, *65*, 24–32. <https://doi.org/10.1007/s11881-015-0098-7>.
 17. Shalileh, S.; Ignatov, D.; Lopukhina, A.; Dragoy, O. Identifying Dyslexia in School Pupils from Eye Movement and Demographic Data Using Artificial Intelligence. *PLOS ONE* **2023**. <https://doi.org/10.1371/journal.pone.0292047>.
 18. Chowdhury, M.S. Comparison of accuracy and reliability of random forest, support vector machine, artificial neural network and maximum likelihood method in land use/cover classification of urban setting. *Environmental Challenges* **2024**, *14*, 100800. <https://doi.org/https://doi.org/10.1016/j.envc.2023.100800>.
 19. Ngo, V.M.; Helmer, S.; Le-Khac, N.A.; Kechadi, M.T. Structural textile pattern recognition and processing based on hypergraphs. *Information Retrieval Journal* **2021**, *24*, 137–173. <https://doi.org/10.1007/s10791-020-09384-y>.
 20. Kamari, A.; Peter Leslie Schultz, C. A combined principal component analysis and clustering approach for exploring enormous renovation design spaces. *Journal of Building Engineering* **2022**, *48*, 103971. <https://doi.org/https://doi.org/10.1016/j.jobbe.2021.103971>.
 21. Department of Education. *Guidelines for Primary Schools Supporting Pupils with Special Educational Needs in Mainstream Schools*; Ireland Government, 2021.
 22. Ngo, V.M.; Cao, T.H. A Generalized Vector Space Model for Ontology-Based Information Retrieval. *Vietnamese Journal on Information Technologies and Communications* **2009**, *22*, 43–53.
 23. Dang, C.; Moreno-García, M.N.; Prieta, F.; Nguyen, K.V.; Ngo, V.M. Sentiment Analysis for Vietnamese – Based on Hybrid Deep Learning Modelse. In Proceedings of the Proceedings of the 18th International Conference on Hybrid Artificial Intelligence Systems (HAIS 2023), 2023, Vol. 14001, *LNAI*, pp. 293–303. https://doi.org/10.1007/978-3-031-40725-3_25.
 24. Ngo, V.M.; Duong, T.V.T.; Nguyen, T.B.T.; Nguyen, P.T.; Conlan, O. An Efficient Classification Algorithm for Traditional Textile Patterns from Different Cultures Based on Structures. *Journal on Computing and Cultural Heritage (JOCCH)* **2021**, *14*, 1–22.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.