CONTENTS

DIAGNOSING DYSLEXIA IN CHILDREN VIA EYE-TRACKING

Anonymous authors
Paper under double-blind review

ABSTRACT

Dyslexia is a neurodevelopmental disorder that impairs reading ability without affecting general intelligence. Early detection is critical for initiating timely rehabilitation and improving academic outcomes. However, conventional diagnostic methods are often time-consuming, subjective, and dependent on expert evaluation. This study presents a novel dyslexia detection approach using eye-tracking data from two reading tasks, processed through a deep learning model based on Operational Neural Networks (ONNs). The model utilizes three input types: gaze coordinates from both tasks and nine cognitively-informed spatiotemporal and efficiency-based gaze features. It achieved a classification accuracy of 93.55% and an F1-score of 0.9375, surpassing several traditional machine learning models. Our findings show that integrating raw gaze data with a compact, meaningful feature set enables ONNs to effectively distinguish dyslexic from non-dyslexic readers. This approach offers a fast, objective, and scalable tool for early dyslexia screening, with strong potential for educational applications. By supporting early and accurate diagnosis, it contributes to timely, personalized intervention strategies, which are crucial for mitigating the long-term impact of dyslexia.

1 INTRODUCTION

Dyslexia is a neurobiological learning disability characterized by persistent difficulties in learning to read. One of the most commonly referenced definitions of dyslexia originates from a collaborative effort among major organizations in the field, including the International Dyslexia Association, the National Center for Learning Disabilities, and the National Institute for Child Health and Human Development [Lyon et al. (2003)].

> “Dyslexia is a specific learning disability that is neurobiological in origin. It is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language that is often unexpected in relation to other cognitive abilities and the provision of effective classroom instruction. Secondary consequences may include problems in reading comprehension and reduced reading experience that can impede growth of vocabulary and background knowledge. (p. 2)”

Dyslexia is commonly estimated to affect approximately 3% to 7% of the population, based on the criterion of performing 1.5 standard deviations or more below the average on standardized reading assessments.[Peterson & Pennington (2012); Snowling & Melby-Lervåg (2016)]

Early identification is crucial, as undiagnosed dyslexia can negatively impact academic performance and self-esteem. Traditional diagnosis relies on time-consuming reading tests and expert judgment, but eye-tracking technology offers an objective window into reading behavior.

Research over the past two decades has increasingly explored machine learning (ML) models that analyze eye movement patterns to automatically detect dyslexia in children. Eye-tracking can capture telltale signs of dyslexia – such as longer fixations, more frequent regressions (backward eye movements), and irregular saccades in real time, providing rich data for ML-based classification. This review synthesizes findings from the last 20 years (2005–2025) on dyslexia diagnosis in children using ML and eye-tracking data. We cover peer-reviewed studies and notable preprints, highlighting key trends, features, algorithms, data sets, and diagnostic performance (accuracy, sensitivity, etc.). We also critically compare methodologies, including classical approaches (e.g., support vector machines, decision trees) and newer deep learning techniques, to assess the state of the art in this emerging interdisciplinary field.

Extensive research confirms that dyslexic readers exhibit distinctive eye movement patterns during reading. Compared to typical readers, children with dyslexia tend to have longer and more frequent fixations, shorter saccades, and a higher number of regressions. For example, dyslexic readers often fixate on words 200–300 ms or longer and frequently re-read text, whereas fluent readers make quicker, more linear progress. These differences reflect the greater cognitive effort dyslexic children expend in decoding. Such metrics fixation duration, fixation count, saccade length, regression count – have been identified as diagnostic features for dyslexia. A dyslexic child's gaze path shows numerous backtracking (regressive) saccades and prolonged dwells on words (1). By contrast, a typical reader's path would show shorter, evenly paced fixations moving left-to-right (2). These eye movement disparities form the basis for ML models to classify readers as dyslexic or not.

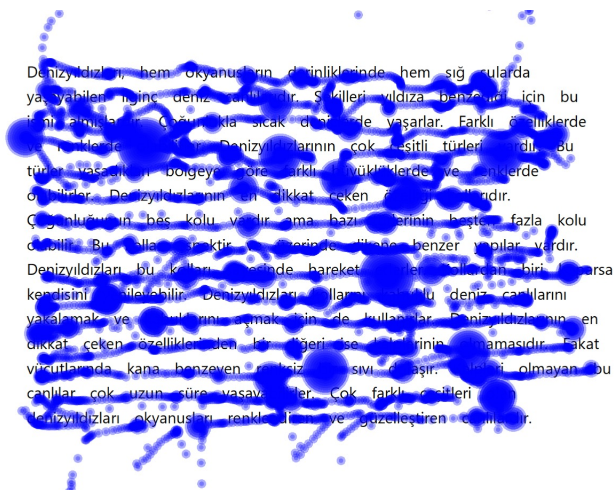

Figure 1: Example Reading of Dylexic Child On the Didactic Text

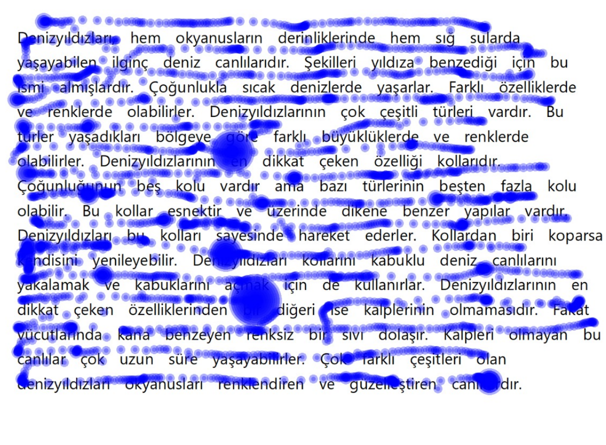

Figure 2: Example Reading of Typical Child On the Didactic Text

2 RELATED WORK

Early efforts to automatically classify dyslexia from eye-tracking data applied traditional machine learning with engineered features. Rello & Ballesteros (2015) conducted one of the first studies, extracting simple features like the number and duration of fixations and visits (re-reading instances) from eye-tracking of 97 Spanish readers (48 with dyslexia, ages 11–54). Using a support vector machine (SVM) classifier on these features, their model achieved 80% accuracy in distinguishing dyslexic vs. non-dyslexic readers. This foundational work demonstrated the feasibility of ML-based dyslexia screening, albeit with moderate accuracy given the broad age range and limited feature set. Subsequent studies focused on children in narrower age ranges and expanded feature engineering to improve performance.

A landmark study by Ekstrand et al. (2021) used a comprehensive feature set and rigorous feature selection to push accuracy into the mid-90% range. They collected eye movements from 185 Swedish 9–10 year-olds (97 identified as high-risk for dyslexia) during a one-minute reading test. From each recording, 168 quantitative features describing fixations and saccades (e.g. durations, amplitudes, variability, spatial distribution) were extracted. A recursive feature elimination procedure (SVM-RFE) identified an optimal subset of 48 features for classification. Using a linear SVM with nested cross-validation, the model reached a best accuracy of 95.6%, with balanced sensitivity and specificity around 95% . This high accuracy – achieved in under one minute of eye-tracking – was a proof-of-concept that dyslexia risk can be detected objectively with eye movement data at a performance level comparable to or better than standard reading tests. Notably, the most predictive features included average fixation duration, number of fixations, and number of backward (regressive) saccades, aligning with known dyslexic reading behaviors.

Several studies have since replicated and built upon these results. Appadurai & Bhargavi (2019) re-analyzed the same Swedish dataset with various ML algorithms, confirming that both SVM and tree-based classifiers can attain 95% accuracy. For instance, Prabha et al. report that k-nearest neighbors and hybrid SVM models with particle swarm optimization (for feature weighting) reached 95–96% accuracy on the task Their work also explored reducing the feature set to only fixation-related metrics without significant loss of accuracy.

Similarly, Asvestopoulou et al. (2019) applied SVM, Naïve Bayes, and clustering methods to a Greek child eye-tracking dataset (69 children, age 8.5–12.5. Intriguingly, they found that a minimal feature subset could suffice: using just three features – saccade length, count of short forward movements, and count of repeatedly fixated words – their classifier achieved 97% accuracy. This suggests that a few well-chosen eye metrics can be highly discriminative of dyslexia, at least within a controlled setting. However, generalizing feature importance across languages and texts remains challenging; for example, fixation time might be a stronger indicator in languages with complex orthography.

Beyond static reading tests, some researchers have incorporated varied tasks and broader age ranges. Andreadakis et al. (2017) introduced a rapid screening method (RADAR) for Greek children using two reading passages of different difficulty. A threshold-based algorithm (TRS) on features like fixation count and mean fixation duration yielded 94.2% accuracy on an age-appropriate text, dropping to 88% on an easier text for younger readers . This drop highlights that dyslexic eye movement patterns become less pronounced with simpler material.

El Hmimdi et al. (2021) took a multi-task approach with French adolescents, combining one reading test and two non-reading gaze tasks (visual exploration) using a wearable eye-tracker. Using a variety of features and ML models, they achieved 81% accuracy on both reading and non-reading tasks. Notably, their best performing model had 81.3% sensitivity for dyslexia and similar specificity, indicating a modest trade-off in recall compared to the highly controlled reading-only studies. These multi-modal results underscore that while reading eye movements are most directly relevant for dyslexia, certain oculomotor patterns in non-reading tasks might also carry diagnostic signal (though at lower accuracy).

In recent years, deep learning approaches have emerged to classify dyslexia from raw eye-tracking data, aiming to bypass manual feature engineering. Nerusil et al. (2021) proposed a “holistic” method that feeds entire eye movement sequences into a convolutional neural network (CNN). They used the same large Swedish dataset (185 children) but applied minimal preprocessing: gaze coordinate sequences were interpolated to a fixed length and transformed (e.g. via a Fourier-based magnitude spectrum) for input to a 1D CNN. The CNN achieved 96.6% accuracy, slightly exceeding the best SVM-RFE result on that dataset (95.6%). The authors note this as evidence that a deep learning model can automatically extract predictive spatio-temporal features, matching the performance of meticulously crafted features.

Likewise, ***a 2022 study by Vajs et al.*** encoded children's eye-tracking data as color-coded images of gaze trajectories and trained a deep CNN, obtaining 87% accuracy. However, deep learning models typically require larger training sets and careful validation to avoid overfitting. When researchers attempted cross-dataset evaluation, performance often dipped (e.g. 83–86% in cross-language tests), highlighting that models tuned on one dataset may not directly transfer to another without some loss. **[[BULAMADIM, 2023 OCAK BULDUM, SONUÇLARI : The best-achieved accuracies were 85.6% when evaluated on DS1 and 82.9% when evaluated on DS2. : 10.1109/ACCESS.2023.3234438]]**

Across studies, a core set of eye movement features consistently emerges as important for dyslexia detection. These include fixation-based metrics (count of fixations, mean and total fixation duration, fixation dispersion), saccade metrics (saccade count, average amplitude/length, frequency of very short saccades, number of regressions), and reading pattern indices (e.g. percentage of words with multiple fixations, ratio of regression to forward saccades). For instance, longer average fixation time and a higher fraction of regressions are strong indicators of dyslexic reading. Some studies incorporated additional features such as variability measures (standard deviation of fixation durations or saccade amplitudes) and spatial metrics (e.g. the distance between consecutive fixations).

Nilsson Benfatto et al. (2016) categorized their 168 features into groups capturing duration, amplitude, directionality, stability, and binocular coordination of eye movements). Subsequent feature ablation analyses found that durations (especially fixation duration) and frequency counts (number of fixations/regressions) carried the most predictive power. On the other hand, certain complex features like binocular disparity measures were pruned out by feature selection as they contributed little. Using SVM with recursive feature elimination (SVM-RFE), they identified an optimal subset of 48 features, reducing the feature space by 71% while achieving 95.6% ± 4.5% accuracy with balanced sensitivity (95.5%) and specificity (95.7%). While randomly selected features (n=126) achieved comparable accuracy (95.3%), they resulted in a more complex model with only a 25% feature reduction. Notably, chance-level models performed significantly worse, confirming the reliability of the selected features

A notable finding by Raatikainen et al. (2021) was that features derived from transition matrices – describing probabilities of moving between lines or words – can be useful: their hybrid Random Forest + SVM model identified such features as important and attained 89.7% accuracy with a recall of 85%. In general, combining complementary features (fixation and saccade metrics together) yields better accuracy than any single type alone. Yet, as mentioned, even a few optimized features can achieve high accuracy on specific datasets. This has practical implications: simpler feature sets could allow faster, more interpretable screening tools, though possibly at a cost of universality.

On the ML side, a wide spectrum of algorithms has been employed. Support Vector Machines (SVMs) have been most common in early works due to their effectiveness on high-dimensional data (Nilsson Benfatto et al. (2016)). SVMs (often with linear kernels) yielded strong results (80–96% accuracy) in Rello & Ballesteros (2015), Nilsson Benfatto et al. (2016), and many follow-up studies (Švaříček et al. (2025))(Nilsson Benfatto et al. (2016)). Decision tree ensembles like Random Forests have also been used, either alone or to rank features for another classifier. For instance, Raatikainen et al. (2021) used a Random Forest to select top features, then fed those into an SVM. In some cases, k-nearest neighbors (kNN) performed as well as SVM – one study reported kNN slightly outperforming SVM and Random Forest (95% vs 92%) on a fixation-feature set. Simpler methods like logistic regression and Naïve Bayes have generally shown slightly lower performance **(e.g. Vajs et al. noted logistic regression reached 90% vs 94% for SVM/kNN on their data)**. Deep neural networks (CNNs) and even recurrent networks have gained traction more recently, as discussed. CNNs have been applied both to raw gaze sequences (1D CNN) and to image-like representations of eye data (2D CNN). The deep models can capture complex temporal patterns that fixed statistical features might miss, but require more data and careful validation. A noteworthy point is that classical ML with well-crafted features has remained competitive with deep learning in terms of pure accuracy on in-sample tests – e.g. both SVM with RFE and a CNN achieved 95–96% on the same data. Thus, no single algorithm universally dominates; rather, success depends on the feature quality, size of training set, and consistency of the evaluation protocol.

In this study, we propose a novel approach for dyslexia detection based on eye movement patterns, leveraging deep learning techniques to analyze gaze behavior during reading tasks. While traditional diagnostic methods for dyslexia often rely on extensive psychoeducational assessments, they can be time-consuming, subjective, and inaccessible in many contexts. Eye tracking, by contrast, provides a non-invasive, objective window into the cognitive processes involved in reading, particularly fixations, saccades, and regressions—features frequently altered in individuals with dyslexia.

By systematically modeling these patterns, we aim to identify reliable biomarkers of dyslexic reading behavior. Our approach utilizes a comprehensive dataset collected from participants with and without dyslexia during controlled reading experiments, extracting spatiotemporal features from their eye movements. These features are then fed into a operationalneural network classifier trained to distinguish dyslexic from non-dyslexic readers. The ultimate goal of our work is to offer a fast, accessible, and accurate tool for early dyslexia screening, which can be integrated into educational or clinical settings. By combining advancements in eye-tracking technology with deep learning, our method contributes to the growing body of research focused on digital, data-driven cognitive diagnostics. In the following sections, we detail the dataset, model architecture, experimental setup, results, and implications for future applications in assistive technology.

3 METHOD

This study utilized a proprietary eye-tracking dataset developed in collaboration with Eyesoft, a company specializing in eye-tracking systems. The participant cohort consisted of children enrolled in the second and third grades of primary school, recruited from local educational institutions following approval from institutional review boards and with informed consent obtained from parents or legal guardians. Each child completed a standardized reading task composed of two textual stimuli: a didactic text designed to simulate textbook-like reading material and a narrative passage intended to capture naturalistic story reading behavior. These texts were presented in a digital format on a screen equipped with a remote, high-speed eye-tracking system capable of capturing gaze data at both high temporal (minimum 120 Hz sampling rate) and spatial precision (error margin below 1° of visual angle). The system recorded raw gaze coordinates (X and Y positions on the screen), timestamps, fixation durations, and saccadic events throughout the reading tasks.

3.1 COLLECTED DATA

The final dataset comprised eye-tracking recordings from a total of 163 children (83 female, 80 male), with 114 in third grade and 49 in second grade. Based on standardized psychoeducational evaluations conducted by licensed experts, 83 participants were classified as typically developing readers and 80 as likely dyslexic.

Each child completed two reading tasks: one didactic and one narrative, resulting in a total of 326 reading trials.

Each reading trial consisted of:

- A continuous sequence of raw gaze coordinates (X, Y) captured at $\geq$120 Hz temporal resolution during the task;
- A corresponding vector of nine spatiotemporal and efficiency-based engineered features, computed from the raw gaze coordinate sequences (X, Y) recorded during each trial;
- A binary class label indicating dyslexia status (1 = dyslexic, 0 = non-dyslexic).

All input sequences were normalized and either truncated or zero-padded to a fixed temporal length, ensuring consistent input shape for the deep learning architecture. This design allowed for the effective fusion of high-resolution sequential data and cognitively-informed summary features within a unified classification framework

3.1.1 PREPROCESSING

Raw eye-tracking data underwent a rigorous preprocessing pipeline to enhance signal clarity and ensure consistency across samples. The initial stage of preprocessing involved the removal of biologically implausible data points, including coordinates outside screen boundaries, zero-valued samples, or prolonged blinks misclassified as fixations. Following this initial cleaning, gaze trajectories were smoothed using a derivative-based thresholding technique to eliminate high-frequency noise, reduce micro-saccadic jitter, and ensure that the transitions between fixations and saccades were appropriately delineated. Once gaze data were cleaned, each participant's reading session was segmented according to the two distinct texts, and behavioral features were extracted independently for each session to capture task-specific eye movement patterns.

From each reading trial, a fixed set of nine quantitative features was derived based on both prior literature and empirical exploration of feature utility in distinguishing dyslexic and non-dyslexic reading behaviors. These features included: total time spent reading the passage; the standard deviation of gaze points in the horizontal and vertical dimensions, serving as proxies for visual scanning breadth and reading line consistency; the number of regressions, operationalized as saccades directed backwards against the reading flow; the maximum, minimum, and mean saccade lengths in pixels; and a temporal efficiency measure denoting the time required for the participant to accumulate 30 seconds of active reading time. These features were chosen to encapsulate spatiotemporal aspects of visual attention allocation and control, which are often atypical in children with dyslexia.

To ensure robustness and minimize the influence of outliers, we applied the Interquartile Range (IQR) method to all continuous features. Data samples with any feature value falling outside 1.5 times the IQR from the first or third quartile were flagged as anomalies and excluded from the analysis. This step was critical to prevent overfitting and to ensure that extreme behaviors, potentially due to distractions or equipment error, did not bias the training process. Moreover, because deep learning architectures require inputs of uniform dimensionality, all gaze-derived sequences were truncated or zero-padded to a consistent sequence length, preserving temporal order and ensuring compatibility with batch-based model training.

3.1.2 LABELING AND DATASET CONSTRUCTION

Each participant's classification label was obtained through an independent diagnostic assessment carried out by licensed educational psychologists affiliated with the participating schools. These assessments were based on standardized psychoeducational testing protocols evaluating phonological processing, reading fluency, and decoding skills, which collectively inform the clinical judgment of whether a child is at risk for, or exhibits, dyslexia. For modeling purposes, the outcomes of these expert evaluations were binarized into two distinct classes: "likely dyslexic" and "non-dyslexic." The final dataset thus consisted of paired input sequences (eye-tracking features from both text types) and binary labels for each participant, forming the basis of a supervised classification problem. [CEVRİYE HOCA]

3.2 MODEL

3.2.1 MODEL ARCHITECTURE

The model receives three distinct input streams: (1) a preprocessed sequence of 2D gaze coordinates (x, y) from the first reading task as mentioned in section 3.1.1, (2) a similar sequence from the second reading task, and (3) a numerical feature vector summarizing each student's reading behavior. This vector consists of 9 features extracted from the full eye-tracking session: total reading time, horizontal and vertical gaze deviation, number of regressions (backward saccades), mean and maximum saccade distances, standard deviation of x-coordinates, and number of words read within 30 seconds.

To leverage the sequential and nonlinear nature of eye movement data, we implemented a custom deep learning model based on the Operational Neural Network (ONN) architecture. Unlike conventional Convolutional Neural Networks (CNNs) which employ fixed mathematical operations (typically inner product and summation) to perform feature extraction, ONNs extend this approach by learning a diverse set of non-linear operators for each layer. This flexibility allows ONNs to adapt their behavior at the neuron level, enabling a more expressive and dynamic feature learning process well-suited to complex cognitive patterns such as reading behavior.[Kiranyaz et al. (2019)]

The network was structured with multiple operational convolutional layers, each followed by batch normalization and a non-linear activation function selected from a pool of parametric operators. These were designed to preserve spatial dependencies within the gaze data while also capturing higher-order interactions between features. Dropout regularization was employed after each convolutional block to reduce the risk of overfitting, especially given the modest sample size relative to model complexity. The architecture concluded with a fully connected layer that aggregated feature representations across time, feeding into a final sigmoid activation unit to yield binary classification outputs. The entire model was trained in an end-to-end fashion, allowing both low-level feature learning and high-level decision-making layers to be jointly optimized.

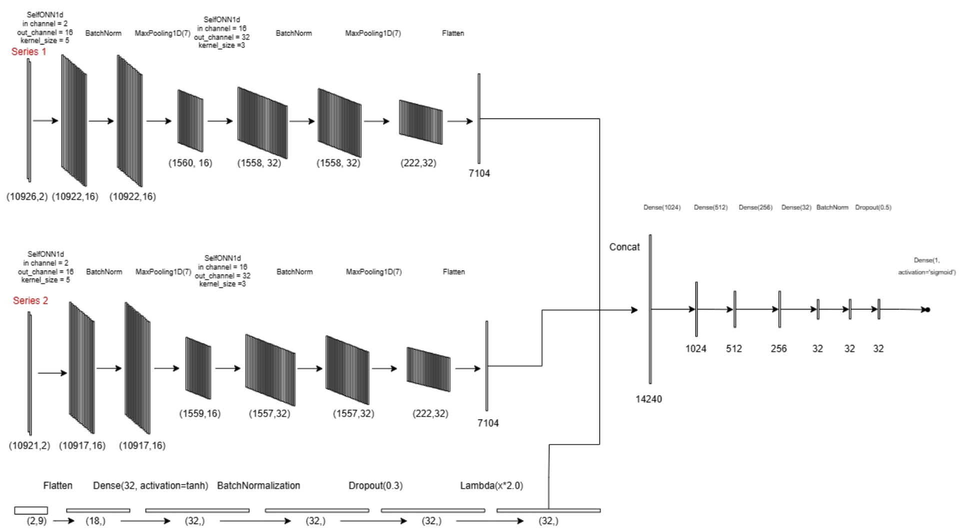

Figure 3: Architecture of the model used

3.2.2 Training Procedure

The full dataset was evaluated using stratified 5-fold cross-validation, ensuring that each fold maintained proportional representation of both classes to prevent class imbalance. At each iteration, four folds were used for training and one for validation/testing, allowing for a robust assessment of model generalizability across different data subsets. During training, the model was optimized using the Adam optimizer with binary cross-entropy loss as the objective function. Hyperparameters such as the learning rate, dropout rate, batch size, and the number of epochs were optimized through grid search in conjunction with five-fold cross-validation on the training and validation sets. Early stopping was implemented to monitor validation loss and prevent overfitting by halting training once performance ceased to improve over a specified number of epochs.

Model performance was quantitatively assessed using a comprehensive suite of classification metrics that included overall accuracy, precision, recall (sensitivity), specificity, and F1 score. The final ONN model achieved a classification accuracy of 93.55% on the held-out test set, with a corresponding F1-score of 0.9375. This result indicates a strong generalization ability and balanced predictive performancinoth classes. In particular, the high F1-score suggests that the model maintained an appropriate balance between precision and recall, which is especially important in educational diagnostics, where both false positives and false negatives can have significant consequences.

In sum, the methodological pipeline described above—from high-resolution data collection and careful preprocessing to the deployment of a robust, interpretable deep learning framework—demonstrates the feasibility of detecting dyslexia in children through eye-tracking data. The combination of operational neural networks and thoughtfully engineered input features represents a promising direction for scalable, non-invasive, and objective screening tools for early reading difficulties.

4 Results and Discussion

4.1 Model Performance

The proposed Operational Neural Network (ONN) model achieved a classification accuracy of 93.55% on the held-out test set, along with an F1-score of 0.9375, indicating a strong balance between precision and recall. These results position the ONN approach as a competitive and robust alternative to traditional machine learning models such as SVMs and Random Forests, which have typically reported accuracies in the 80–95% range across similar datasets. The high F1-score is particularly noteworthy in the context of educational diagnostics, where both false positives (misclassifying non-dyslexic children) and false negatives (failing to detect dyslexia) can carry significant consequences.

Compared to prior deep learning approaches that relied on CNNs applied to either raw gaze sequences or gaze trajectory images, our model demonstrates comparable or superior performance, while also benefiting from the flexibility of learning nonlinear operators tailored to the unique spatiotemporal patterns in eye-tracking data. This result supports the hypothesis that dyslexia manifests in measurable, learnable patterns of visual attention during reading, and that advanced architectures like ONNs are well-suited to capture such complexity.

4.2 Feature Analysis and Interpretation

Although the ONN model operates in an end-to-end fashion, our input features were informed by established markers of dyslexic reading behavior, such as increased fixation duration, higher regression counts, and shorter saccade lengths. Feature selection was driven not only by theoretical relevance but also by empirical performance, ensuring that the model learned from cognitively meaningful data. The inclusion of both spatiotemporal and efficiency-based metrics (e.g., time to reach 30 seconds of active reading) allowed the model to account for both spatial irregularities in gaze and processing speed, both of which are commonly impaired in dyslexic readers. Interestingly, the relatively small feature set—comprising only nine derived metrics—was sufficient to yield high accuracy, reinforcing findings from prior studies (e.g., Asvestopoulou et al. (2019)) that even minimal but well-selected features can effectively discriminate between dyslexic and non-dyslexic reading patterns. This has practical implications for scalable deployment: a lightweight feature set can reduce preprocessing time and model complexity without sacrificing diagnostic utility.

4.3 Limitations and Future Directions

Despite promising results, several limitations must be acknowledged. First, the dataset was limited to a specific age group (second and third graders) and a particular language context, which may constrain the generalizability of the model across broader populations and orthographic systems. Cross-linguistic validation is a critical next step, as eye movement patterns can vary based on the transparency and complexity of written language. Second, while the ONN architecture offers strong performance, its interpretability remains a challenge. Future work could incorporate explainable AI (XAI) techniques to better understand which features or gaze behaviors most heavily influence the model's decisions. Additionally, expanding the dataset to include longitudinal data could enable the development of models that not only classify dyslexia but also track reading development and intervention response over time.

5 Conclusion

This study presents a novel deep learning approach for detecting dyslexia in children based on eye-tracking data collected during reading tasks. Leveraging the expressive capabilities of Operational Neural Networks, our model achieved high classification performance with a relatively small and interpretable feature set. The results underscore the potential of combining eye-tracking technology with advanced machine learning to develop fast, non-invasive, and objective screening tools for early detection of reading difficulties. By reducing reliance on time-consuming and subjective traditional assessments, such tools can facilitate earlier interventions and improve educational outcomes for children at risk of dyslexia. Future research should aim to validate these findings across diverse linguistic, cultural, and age contexts, and explore integration into classroom or clinical environments as part of broader assistive diagnostic platforms.

6 Acknowledgments

We extend our sincere thanks to Eyseoft for their support during the data collection process. We are grateful to Prof. Dr. Cevriye Ergül for her invaluable contribution in administering standardized assessments used for labeling and for her guidance in preparing the reading materials utilized during data acquisition.

7 Ethics Statement

All procedures involving human participants were conducted in accordance with ethical standards and approved by the Turkish Ministry of National Education (MEB). Informed consent was obtained from the legal guardians of all participating children prior to data collection.

References

Vassilios Andreadakis, Vasileios Selimis, Michail Kalaitzakis, Bachourou Theodora, Georgios Kaloutsakis, George Kymionis, Stelios Smirnakis, and Ioannis Aslanides. Radar: A novel fast-screening method for reading difficulties with special focus on dyslexia. *PLOS ONE*, 12: e0182597, 08 2017. doi: 10.1371/journal.pone.0182597.

Jothi Appadurai and R. Bhargavi. *Prediction of Dyslexia Using Machine Learning—A Research Travelogue*, pp. 23–34. 01 2019. ISBN 978-981-13-7090-8. doi: 10.1007/978-981-13-7091-5_3.

Thomais Asvestopoulou, Victoria Manousaki, Antonis Psistakis, Vassilios Andreadakis, Ioannis Aslanides, and Maria Papadopouli. Dyslexml: Screening tool for dyslexia using machine learning, 03 2019.

Anna Ekstrand, Mattias Nilsson, and Gustaf Öqvist Seimyr. Screening for reading difficulties: Comparing eye tracking outcomes to neuropsychological assessments. *Frontiers in Education*, 6, 03 2021. doi: 10.3389/feduc.2021.643232.

Alae Eddine El Hmimdi, Lindsey M Ward, Themis Palpanas, and Zoï Kapoula. Predicting dyslexia and reading speed in adolescents from eye movements in reading and non-reading tasks: A machine learning approach. *Brain Sciences*, 11(10), 2021. ISSN 2076-3425. doi: 10.3390/brainsci11101337. URL https://www.mdpi.com/2076-3425/11/10/1337.

Serkan Kiranyaz, Turker Ince, Alexandros Iosifidis, and Moncef Gabbouj. Operational neural networks, 02 2019.

G. Lyon, Sally Shaywitz, and Bennett Shaywitz. A definition of dyslexia. *Annals of Dyslexia*, 53: 1–14, 08 2003. doi: 10.1007/s11881-003-0001-9.

Boris Nerusil, J. Polec, Juraj Skunda, and Juraj Kacur. Eye tracking based dyslexia detection using a holistic approach. *Scientific Reports*, 11, 08 2021. doi: 10.1038/s41598-021-95275-1.

Mattias Nilsson Benfatto, Gustaf Öqvist Seimyr, Jan Ygge, Tony Pansell, Agneta Rydberg, and Christer Jacobson. Screening for Dyslexia Using Eye Tracking during Reading. *PLoS ONE*, 11 (12):e0165508, December 2016. doi: 10.1371/journal.pone.0165508.

Robin L Peterson and Bruce F Pennington. Developmental dyslexia. *The Lancet*, 379 (9830):1997–2007, 2012. ISSN 0140-6736. doi: https://doi.org/10.1016/S0140-6736(12)60198-6. URL https://www.sciencedirect.com/science/article/pii/S0140673612601986.

Peter Raatikainen, Jarkko Hautala, Otto Loberg, Tommi Kärkkäinen, Paavo Leppänen, and Paavo Nieminen. Detection of developmental dyslexia with machine learning using eye movement data. *Array*, 12:100087, 2021. ISSN 2590-0056. doi: https://doi.org/10.1016/j.array.

2021.100087. URL https://www.sciencedirect.com/science/article/pii/S2590005621000345.

Luz Rello and Miguel Ballesteros. Detecting readers with dyslexia using machine learning with eye tracking measures. pp. 1–8, 05 2015. doi: 10.1145/2745555.2746644.

Margaret Snowling and Monica Melby-Lervåg. Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological bulletin*, 142, 01 2016. doi: 10.1037/bul0000037.

Roman Švaříček, Nicol Dostalova, Jan Sedmidubsky, and Andrej Cernek. Insight : Combining fixation visualisations and residual neural networks for dyslexia classification from eye-tracking data. *Dyslexia*, 31, 01 2025. doi: 10.1002/dys.1801.