Dyslexia Analysis and Diagnosis Based on Eye Movement

R. Vaitheeshwari, Chih-Hsuan Chen, Chia-Ru Chung, Hsuan-Yu Yang, Shih-Ching Yeh, Eric Hsiao-Kuang Wu, *Member, IEEE*, and Mukul Kumar

***Abstract*—Dyslexia is a complex reading disorder characterized by difficulties in accurate or fluent word recognition, poor spelling, and decoding abilities. These challenges are not due to intellectual, visual, or auditory deficits. The diagnosis of dyslexia is further complicated by symptom variability, influenced by cultural and personal factors. This study leverages Virtual Reality (VR) advancements, eye movement tracking, and machine learning to create a virtual reading environment that captures eye movement data. This data extracts features such as eye movement metrics, word vectors, and saliency maps. We introduce a novel fusion model that integrates various machine learning algorithms to objectively and automatically assess dyslexia using physiological data derived from user interactions. Our findings suggest that this model significantly enhances the accuracy and efficiency of dyslexia diagnosis, marking an important advancement in educational technology and providing robust support for individuals with dyslexia. Although the sample size was limited to 10 dyslexic and 4 control participants, the results offer valuable insights and lay the groundwork for future studies with larger cohorts.**

***Index Terms*—Cognitive assessment, diagnostic tools, dyslexia, eye movement tracking, fusion models, machine learning, physiological data analysis.**

I. Introduction

DYSLEXIA is the most common type of learning disability [1], [2]. It is characterized by difficulties in reading and writing despite the absence of intellectual or sensory impairments. Dyslexia usually appears before the age of 8 and often results in low academic achievement by the third grade [3]. In general, an estimated 3% to 7% of the population is affected by dyslexia [4], [5]. In Taiwan, approximately 43,000 people are identified with learning disabilities [6]. Early identification and intervention can significantly reduce the need for special assistance later, thereby decreasing social costs and enhancing the quality of life [7].

Our research is dedicated to enhancing the learning capabilities and overall quality of life for children from diverse linguistic backgrounds. We aim to refine educational methodologies and tools to address the unique challenges posed by linguistic diversity, which significantly impedes the accurate diagnosis of dyslexia. Collaborating with linguists and utilizing datasets from various linguistic environments, our tools are designed to be adaptable to cultural nuances. While our long-term vision includes developing universally applicable solutions, this study focuses on applying advanced machine learning techniques, particularly the Bidirectional Encoder Representation from Transformers (BERT) model integrated with eye-tracking data, to analyze dyslexic reading patterns in Chinese scripts.

A key innovation of our approach is the use of a Virtual Reality (VR) environment, departing from conventional diagnostic methods. This approach is demonstrated through a pilot study with 10 dyslexic and 4 control participants. Our study focuses on addressing the complexities of Chinese scripts and presents initial findings that underscore the potential for larger-scale studies and the adaptation of our methods across different languages and cultural contexts.

Diagnosing dyslexia is a considerable challenge, involving the identification of students with low academic performance and the subsequent allocation of specialized remedial education. Traditional diagnostic approaches, such as vocabulary judgment, response times, and accuracy assessments [8], [9] often fail to capture the full complexity of dyslexia across different languages and etiologies. This limitation is particularly pronounced in ideographic scripts like Chinese, where dyslexia can manifest differently than in alphabetic languages. For instance, challenges in Chinese scripts include word formation difficulties and visual perceptual disturbances caused by character adjacency and component assembly, significantly affecting reading efficiency.

Our research introduces innovative approaches to overcome these challenges by integrating the BERT model with eye-tracking technology, which significantly enhances diagnostic capabilities within the context of Chinese scripts [10]. This

Received 5 March 2024; revised 18 June 2024 and 10 October 2024; accepted 4 November 2024. Date of publication 11 November 2024; date of current version 22 November 2024. *(Corresponding author: Eric Hsiao-Kuang Wu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the National Cheng Kung University Human Research Ethics Committee under Application No. NCKU HREC-E-108-501-2.

R. Vaitheeshwari, Chia-Ru Chung, Hsuan-Yu Yang, Shih-Ching Yeh, Eric Hsiao-Kuang Wu, and Mukul Kumar are with the Department of Computer Science and Information Engineering, National Central University, Taoyuan 32001, Taiwan (e-mail: vaitheeshwarir@gmail.com; jjrchris@g.ncu.edu.tw; jakeqwaszx77@gmail.com; shihching.yeh@g.ncu.edu.tw; hsiao@csie.ncu.edu.tw; mukulguthalyan@gmail.com).

Chih-Hsuan Chen is with the Department of Special Education, National Taitung University, Taitung 95092, Taiwan (e-mail: ta895105@nttu.edu.tw).

Digital Object Identifier 10.1109/TNSRE.2024.3496087

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

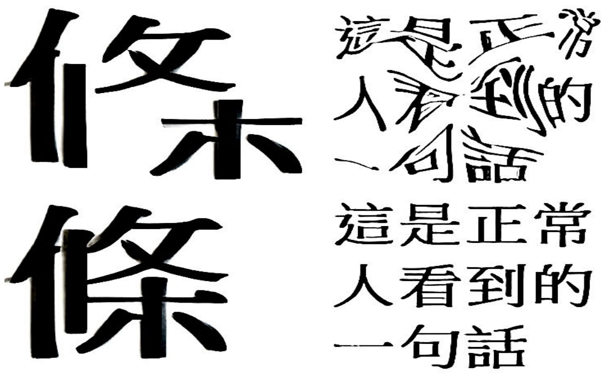

Fig. 1. Comparative visual simulation: dyslexic vs. normal reading perception.

methodology not only advances traditional diagnostic tools but also addresses the specific visual and cognitive challenges presented by Chinese characters.

Figure 1 provides a comparative visual simulation of how dyslexic and non-dyslexic individuals perceive text. The Chinese text translates to 'this is a sentence seen by normal people'. On the top, the characters appear distorted as they might be seen by a dyslexic individual, while the bottom side displays the text as seen by a non-dyslexic individual. Misalignment and incorrect spacing can lead to significant alterations in meanings—for example, the character (tiao) meaning 'strip' could be misinterpreted as (diao) meaning 'drop,' due to visual disturbances. These misinterpretations can drastically change the context and understanding of the text. This visual confusion illustrates how small font sizes and narrow line spacing exacerbate reading difficulties in dyslexic individuals. This visual confusion underscores the specific diagnostic challenges in ideographic scripts and the necessity for objective diagnostic tools [11], [12], [13].

Our VR-based full-article reading analysis has shown significant advantages in diagnosing dyslexia in Chinese scripts, where visual perception plays a key role. Preliminary tests and pilot studies indicate that the VR tools enhance diagnostic precision by dynamically adjusting font size and line spacing. This approach deepens our understanding of dyslexia diagnosis across different linguistic contexts and highlights the potential of adapting our VR-based methodology to other languages. It underscores the need for more nuanced diagnostic approaches to this multifaceted disorder. Beyond its gaming and experiential learning applications, VR offers considerable benefits in educational and research settings, particularly in dyslexia studies. By providing immersive and engaging reading experiences, our approach not only broadens diagnostic capabilities but also improves comfort and engagement, effectively addressing the limitations of standard diagnostic methods.

Additionally, eye movement data plays a crucial role in dyslexia diagnosis. However, traditional eye-tracking methods, which require stationary subjects, are often problematic for young children or those with disabilities [14], [15]. By leveraging a VR reading environment, our study enhances flexibility and control over the diagnostic process, allowing for customizable reading conditions.

Eye-movement metrics such as gaze duration, total reading time, and scanning speed significantly differ between dyslexics and the general population [16], [17]. While electroencephalogram (EEG) [18], [19] and behavioral analyses provide valuable insights, they are often hampered by challenges such as age-related variability and data instability [20]. Our innovative methodology enables a nuanced correlation between quantitative eye-tracking data and qualitative Natural Language Processing (NLP) (BERT) text analysis, thereby enhancing our understanding of dyslexia. By correlating eye movement metrics with the semantic and syntactic features extracted by the BERT model [21], we can infer the cognitive load and processing strategies employed by dyslexic readers. For example, prolonged gaze durations on specific words or phrases may indicate increased cognitive effort required for decoding or comprehension, which is cross-referenced with the linguistic complexity of the text. This approach identifies patterns in how dyslexic individuals process language and provides deeper insights into their reading difficulties. We utilized various machine learning models to analyze the eye-movement data.

We also use saliency maps to identify areas that attract heightened attention from dyslexic readers. Brighter regions indicate greater focus, and when analyzed alongside BERT's semantic assessments, these maps help uncover specific challenges such as word recognition and sentence tracking. The Convolutional Neural Network (CNN) model was utilized to analyse the saliency map as this approach offers deeper insights into the cognitive underpinnings of dyslexia, significantly advancing our diagnostic model.

Our research builds on the diagnostic potential of eye-tracking technology, as highlighted by seminal works such as Dr. Carolien Knoop-van Campen et al.'s study [22], which underscores the value of eye-tracking in understanding dyslexia. Additionally, our study is informed by Yuxin Lin et al.'s research [23], which demonstrates the efficacy of integrating machine learning with eye-tracking data, achieving a 90.06% accuracy rate in depression classification. Similarly, Shivani Choudhary et al. [24] explore how variations in eye movements in response to word changes can shed light on dyslexic reading patterns. These studies collectively support our integrative model, which examines the complex interplay between dyslexia, eye movements, VR, and machine learning. Our research advances the understanding of dyslexic reading patterns through the integration of these technologies and contributes significantly to educational technology and cognitive research. Key contributions of our research include:

- We developed an integrative diagnostic model combining VR, eye movement, and machine learning to enhance dyslexia diagnosis in Chinese scripts.
- We correlated eye-movement data with BERT-extracted semantic features and analyzed saliency maps through CNN to capture both cognitive load and visual reading patterns, improving diagnostic accuracy.
- We implemented a fusion model that combines output from BERT, CNN, and machine learning classifiers through a voting mechanism, effectively reducing false

positives and false negatives for more reliable dyslexia detection.

The manuscript is organized as follows: Section II reviews the literature on dyslexia assessment, eye-movement data analysis, and machine learning, highlighting our novel contributions. Section III outlines the experimental design and methodology in detail. Section IV presents the findings, followed by an in-depth discussion of their implications. Lastly, Section V summarizes the key insights and proposes directions for future research.

II. Related Work

Dyslexia is increasingly recognized as a neurological disorder stemming from complex brain development issues. Current research highlights its correlation with abnormal neural connectivity and auditory processing deficits that affect reading and spelling abilities [25]. Inspired by studies suggesting dyslexia's neural underpinnings [26], [27] our research introduces a novel diagnostic framework that integrates eye-movement analysis specifically tailored to identify various types of dyslexia, including those prevalent in Chinese scripts. This neuroscience-informed approach informs our selection of machine learning features and underpins the application of VR and machine learning technologies, designed to comprehensively assess the impact of dyslexia on reading and cognition.

In revisiting eye-movement experiments for dyslexia diagnosis, our research underscores the significance of employing complete texts, as advocated by Rayner et al. [28]. We advance this methodology through the use of VR, creating a realistic and immersive reading experience that closely replicates natural reading environments. VR allows for the interaction with extensive texts, facilitating a detailed collection of eye-movement metrics such as gaze duration, saccades, and fixations, which are pivotal for analyzing dyslexia, especially within complex scripts like Chinese that traditional methods may inadequately address.

Emerging research corroborates the efficacy of VR in providing insightful assessments, significantly advancing our technological approach and understanding of dyslexia. Jothi Prabha and Bhargavi [29] introduced novel eye-movement analysis features like dispersion threshold and velocity threshold. Our study further explores the effectiveness of velocity-based metrics. Concurrent research by Maria De Luca et al. [30] and Schattka et al. [31] has shown mixed results in the correlation between dyslexia and eye movement patterns, highlighting the need for more refined diagnostic tools. Similarly, Raatikainen et al. [32] applied machine learning to eye-movement data for reading fluency assessments, supported by subsequent statistical analyses by Prabha and Bhargavi [33]. Our study extends these methodologies by employing VR eye-tracking to dissect dyslexia reading patterns, focusing on three primary eye-movement metrics:

- **Sweeping Movements:** We assess saccadic movements frequency, amplitude, and speed to gauge how dyslexic readers navigate texts, providing a baseline for text interaction.
- **Gaze Patterns:** Analyzing fixation durations and locations to identify indicators of reading difficulties prevalent in dyslexia.
- **Word Combinations and Saliency Maps:** Generating these to elucidate reading trajectories and cognitive processing, offering deeper insights into the dyslexic reading experience.

Integrating these metrics with advanced machine learning techniques, such as BERT, aims to refine dyslexia diagnosis by merging technological innovation with psychological and psychoanalytic insights into reading cognition. This interdisciplinary approach enhances our comprehension of dyslexia by correlating eye movement patterns with textual complexity, providing a dual perspective on visual engagement and cognitive processing during reading. In a VR setting, we employ BERT to conduct semantic analysis of the texts read, integrating these findings with eye-tracking data to create a holistic view of how dyslexic individuals process text both visually and linguistically.

Recent BERT advancements have focused on refining the model's architecture, enhancing its applicability across various fields. For example, Eke et al. [34] used BERT for irony detection in social media, achieving high accuracy 98.5%. Similarly, Cai et al. [35] introduced BERT-BiLSTM, combining BERT with BiLSTM for sentiment analysis in the energy sector, reporting an accuracy of 86.2%. In another domain, He and Hu [36] developed a multimodal fusion BERT model that incorporates linguistic, acoustic, and visual data for sentiment analysis, enhancing interaction analysis between different modalities through an innovative update mechanism. Additionally, Lee et al. [37] proposed an approach for multimodal emotion recognition by integrating linguistic, audio, and visual data, setting new performance benchmarks on three datasets.

Our study advances dyslexia research by integrating various machine learning models, including BERT, CNN, and traditional classifiers, to analyze both cognitive and visual patterns. BERT is used to interpret linguistic complexities by analyzing reading behaviors, while CNN processes saliency maps, highlighting key visual indicators such as prolonged fixations and frequent regressions. Traditional machine learning classifiers are applied to analyze eye movement features, including fixation duration and saccade amplitude. By combining these approaches, we capture both cognitive and visual processing anomalies, providing a comprehensive and accurate framework for dyslexia diagnosis.

III. Experimental Setup

A. Participants

In our exploratory study, we analyzed eye movement patterns in dyslexic elementary students using a VR setting. The study included 10 dyslexic students and a control group of 4 non-dyslexic learners, as identified by the Taitung Experimental Association. Table I presents the detailed participant information. While the participant size was limited by feasibility, this pilot study provides preliminary insights into dyslexic reading behaviors. To mitigate biases and ensure accurate

TABLE I
SUBJECT DETAILS

Diagnosis/Statistics	Mean Age	Subject
Normal	10.5	04
Dyslexia	10.1	10

comparison, the non-dyslexic group was referred to as the 'control group,' and language proficiency was standardized across all participants. An orientation session was conducted to familiarize subjects with the VR environment to control novelty effects. This initial exploration highlights the potential for future studies to expand on individual differences such as reading strategies and topic interest.

B. Experimental Equipment

Our experiment utilized a high-fidelity HTC Vive Pro head-mounted display paired with a Tobiivr4 eye-tracking device, ensuring precise data collection. The choice of equipment was predicated on its high-resolution and refresh rate capabilities, pivotal for capturing subtle nuances in eye movement. The monocular resolution is 1440 $\times$ 1600 pixels and the screen update rate is 90Hz.

While the Tobii eye tracker for Vive is designed and marketed for gamers, its specifications are comparable to those of eye trackers commonly used in research literature. For instance, the Tobii VR4 eye tracker offers an accuracy of 0.5 degrees and a sampling rate of 120 Hz, aligning with the standards of widely-used research-grade eye trackers such as the EyeLink 1000 Plus, which also provides an accuracy of 0.5 degrees and a sampling rate of up to 2000 Hz. Although the sampling rate of the Tobii VR4 is lower, its integration with the HTC Vive Pro provides a balanced trade-off between high-quality visual immersion and reliable eye-tracking data. This setup is particularly beneficial in VR environments where a combination of high-resolution displays and accurate eye-tracking is crucial for detailed analysis. Therefore, we believe that the Tobii VR4 eye tracker, despite its gaming origins, is well-suited for our research purposes, providing a robust platform for studying eye movements in a VR setting.

C. Experimental Materials

Reading materials consisted of texts spread across six pages, totalling 574 words (106, 113, 112, 109, 107, and 27 words on each page, respectively). The selection of these texts was carefully controlled for word frequency using standards provided by the Educational Reference Library. Specifically, we employed both high-frequency and low-frequency word dictionaries to ensure a balanced representation of vocabulary that mirrors typical reading experiences of elementary students. High-frequency words were selected based on their regular appearance in elementary-level reading materials, while low-frequency words were chosen for their less common occurrence, providing a comprehensive assessment of reading capabilities across different difficulty levels. This careful selection process aimed to emulate authentic reading experiences, thus ensuring ecological validity.

D. Experimental Process

Prior to formal data collection, subjects underwent several preparatory steps to ensure readiness for the study.

1) **Eye-Tracking Position Correction:** Calibration of the eye-tracker was conducted for each subject to account for individual differences in eye characteristics. This involved an initial calibration procedure using a standard nine-point calibration grid to ensure precise tracking of eye movements.
2) **Eye-Tracking Measures:** We utilized key eye-tracking metrics such as fixation duration, saccade amplitude, and fixation count. Fixation duration refers to the length of time a subject's gaze remains on a single point, saccade amplitude measures the distance between two consecutive fixations, and fixation count indicates the number of fixations within a given reading session. These measures are particularly relevant for identifying dyslexic reading patterns, which often exhibit longer fixations and irregular saccades.
3) **Data Cleaning and Standardization:** Raw eye-tracking data were cleaned to remove any artifacts or noise including filtering out blinks and erratic movements. For data normalization, we applied min-max scaling to adjust for individual variations in reading speed and session length. This ensured that the data from all participants were on a comparable scale. This approach preserves the relationships between data points while accounting for differences in reading pace and session duration, allowing for more accurate cross-subject comparisons. Key metrics like gaze duration and fixation counts were normalized to ensure consistency in analysis.
4) **Descriptive Statistics:** We calculated descriptive statistics for each eye-tracking metric to establish baseline readings and generate hypotheses regarding the differences in eye movement patterns between dyslexic and non-dyslexic learners. This included mean and standard deviation calculations for fixation durations, saccade amplitudes, and fixation counts.
5) **VR Familiarization and Reading Test:** Subjects underwent an orientation session to familiarize themselves with the VR environment, reducing the novelty effect. This was followed by a reading test within the VR setup to assess their engagement and comprehension. The familiarization process included guided tutorials and practice sessions to ensure subjects were comfortable using the VR equipment.
6) **Heat Map Generation:** Heat maps were generated by applying fixation durations and gaze points on the reading material to highlight areas of longer fixation, indicative of greater attention or reading challenges. High sampling rates (90 Hz) were used to capture detailed eye movement data. Smoothing algorithms, such as Gaussian smoothing, were employed to reduce noise and enhance the clarity of the heat maps. Multiple readings per subject were conducted to ensure the reliability of the results. Cross-validation with other

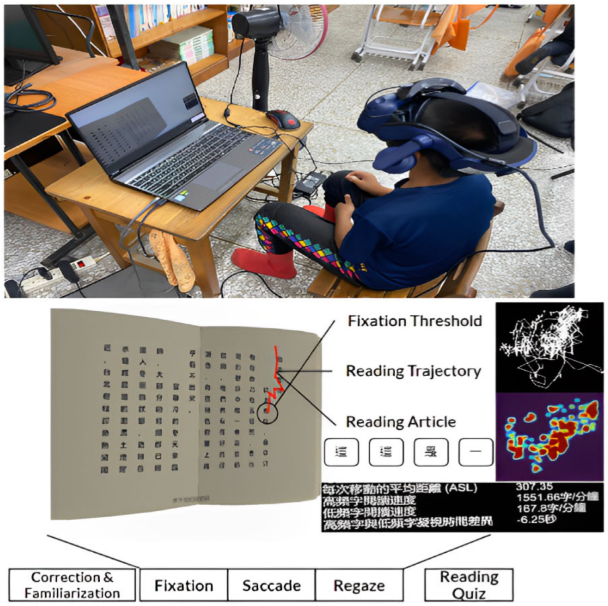

Fig. 2. Eye tracking setup and subject's view during data collection.

eye-tracking metrics was performed to validate the heat maps, and a pilot study was conducted to refine our methodology and ensure robustness.

Figure 2 illustrates our comprehensive eye-tracking setup and data processing workflow. At the top of the figure, we show the participants in the VR setting, where they are engaged in reading tasks. As the student reads, their eye movements are tracked in real-time, and the resulting data (e.g., reading trajectory and fixation points) is overlaid on the reading material, which is depicted in the middle of the figure. The red line indicates the reading trajectory, showing the path the eyes take between fixations, and the circle highlights the "Fixation Threshold," used to determine when fixation is significant. Further along in the figure, the raw eye-tracking data (reading trajectory) is processed into more detailed outputs, such as full trajectory maps (shown on the right in white) and saliency maps (depicted in color). These saliency maps highlight areas of the text that attracted the most visual attention, indicating potential reading difficulties.

At the bottom of the figure, the steps of the overall process are outlined, beginning with calibration and familiarization (labeled "Correction and Familiarization") and continuing through fixation, saccade, and regaze measurements. The final step involves a reading quiz, used to assess the participant's comprehension of the text, connecting the eye-tracking data to cognitive understanding. This workflow demonstrates the flow of information, from capturing the participant's eye movements in the VR environment to generating visual outputs like gaze trajectories and saliency maps, which are then used to evaluate dyslexic reading patterns and comprehension.

E. Machine Learning Model

In this study, we apply advanced machine learning techniques, including BERT and CNN, to tackle the specific challenges of dyslexia diagnosis. By combining these technologies, our approach analyzes complex dyslexic reading patterns, focusing on textual interactions that standard models [38] may overlook. Below, we outline our process, from data collection through to model training and validation for BERT model.

1) **Data Collection:** We extract critical eye-movement features such as fixation duration and saccade patterns after VR reading sessions (as shown in Fig2). These features are known to vary among dyslexic individuals and are essential for identifying dyslexic reading anomalies.
2) **Data Augmentation Techniques:** To enhance the diversity and richness of our dataset, we implement the following augmentation strategies:
   a) **Page Segmentation:** Each reading passage is considered as a separate instance, effectively expanding the size of our dataset.
   b) **Semantic Complexity Adjustment:** We adjust the complexity of text analysis by varying the number of key sentences extracted, which enhances the model's ability to process diverse textual densities. "Key sentences" refer to sentences that carry the most important information, such as those with higher linguistic complexity or crucial for overall comprehension.
   c) **Masking Techniques for Directional Reading:** Masking involves occluding certain portions of the text to simulate natural reading disruptions, such as skipping or rereading words. "Directional reading" refers to the natural flow of reading (e.g., left-to-right or top-to-bottom). This augmentation helps simulate real-world scenarios where reading may be disrupted, thereby helping the model learn to handle these disruptions effectively.
3) **Model Training and Validation:**
   a) **Deep Learning Application:** We utilize both traditional classifiers and advanced deep learning models to analyze the extracted eye movement features. Given the noisy nature of eye-tracking data, deep learning models offer the flexibility needed to customize architectures, accommodating the specific complexities of our data.
   b) **Process Steps:**
      i) **Data Simplification:** We simplify noisy sequences into more manageable data points by focusing on significant dyslexic patterns, such as prolonged fixations and erratic saccades.
      ii) **Sentence Extraction with Snow NLP:** To optimize text processing for Chinese scripts, we use Snow NLP to extract key sentences, which fit within the input constraints of the BERT model. This ensures that the most relevant sentences are analyzed by the model, improving its understanding of reading patterns.
      iii) **Tokenization and Vectorization:** A pre-trained BERT tokenizer transforms these extracted sentences into word vectors,

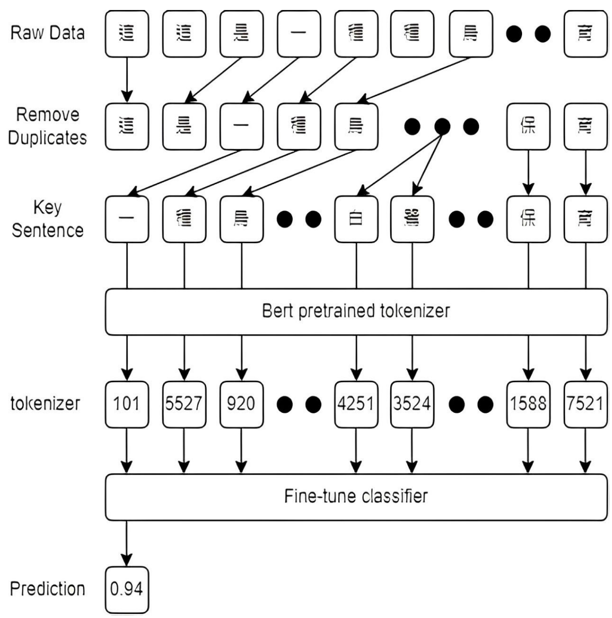

Fig. 3. BERT model detail schematic.

enabling the model to interpret and process the text in a structured format.

iv) **Model Fine-Tuning:** Finally, we fine-tune our BERT model for text analysis and machine learning classifier to improve accuracy in detecting dyslexic reading patterns.

Figure 3 clearly illustrates this flow, depicting the progression from the initial data collection to the preprocessing steps where duplicates are removed and key sentences are selected. The tokenizer then processes these sentences, outputting word vectors that feed into a fine-tuned classifier. This classifier is crucial as it is specifically adjusted to detect patterns indicative of dyslexic reading behaviors. The classifier's adjustments and training are based on the nuances captured from the preprocessed textual data, enabling it to accurately predict dyslexia with high confidence, as represented by the prediction score of 0.94 at the bottom of Fig. 3. Each step in the figure is interconnected, demonstrating the comprehensive analysis from data capture through vectorization to the final prediction, ensuring a clear understanding of the data transformation and the vital role each component plays in the dyslexia detection process.

In our study, we employed advanced visualization techniques to analyze eye movement characteristics, focusing specifically on visual trajectories and saliency maps. Visual trajectories [39] represent the sequential path of gaze movement across text, illustrating how dyslexic and non-dyslexic readers navigate through sentences. These trajectories, plotted using the x and y coordinates of gaze points, help identify patterns such as frequent regressions and prolonged fixations, which are characteristic of dyslexic reading behavior. Fig. 4 illustrates the gaze trajectories of normal (left) and dyslexic (right) readers, showing how each group navigates through text. From bottom to top, the figure displays the sequential path of eye movements, followed by three different types of heat maps generated from these trajectories: a sigmoidal

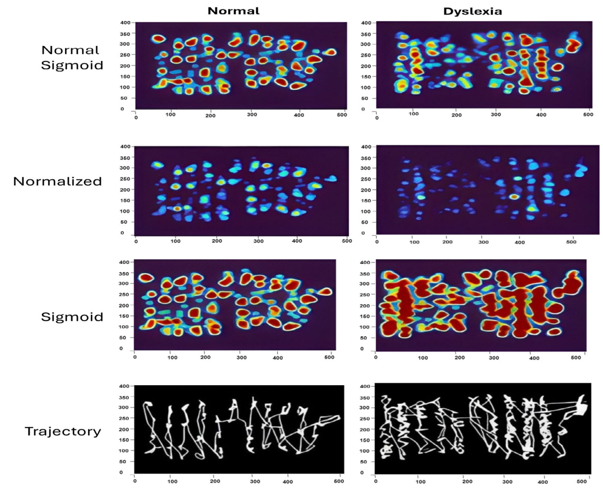

Fig. 4. Eye movement trajectory visualizations in normal vs. dyslexic subjects.

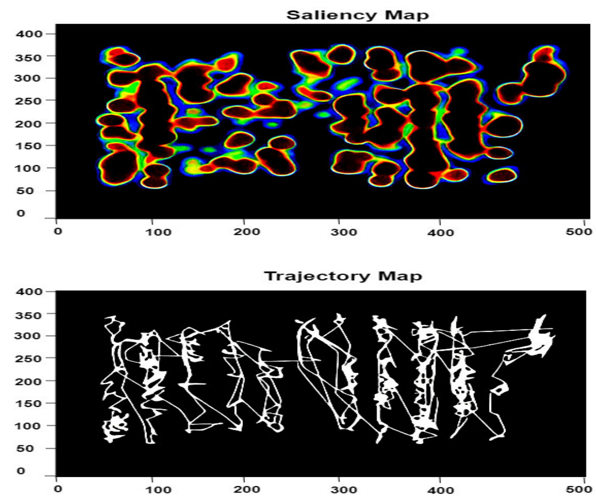

Fig. 5. Eye movement saliency map and trajectory map.

heat map, which emphasizes areas with more gradual changes in gaze density; a normalized heat map, which scales the gaze data uniformly across the entire text; and a combination of sigmoidal and normalized mapping, which balances between gradual gaze transitions and overall gaze distribution. These visualizations help highlight key differences in reading behavior, such as frequent regressions and prolonged fixations, which are characteristic of dyslexic readers.

We also generated saliency maps [40], which overlay heat maps on the text based on the density and duration of gaze points. These maps highlight areas where readers spend more time, indicating sections of text that present reading challenges. The creation of these maps involves calculating the gradient of normalized scores across image pixels, helping to quantify the significance of score changes at a pixel level. Fig. 5 illustrates the gradients used to create these detailed saliency maps, underscoring the precision of our data capture and analysis methods.

The data from visual trajectories and saliency maps are then input into a CNN model for feature extraction. This process involves max-pooling and hidden layer classification to distill the essential features from complex patterns of eye movement. The CNN effectively learns to recognize and categorize these patterns, aiding in the differentiation of dyslexic and non-dyslexic reading behaviors. Fig. 6 illustrates the CNN architecture and highlights how it processes these complex data patterns to produce categorizations, demonstrating the critical role this model plays in our diagnostic tool.

To ensure the robustness and accuracy of our model, we employ a decision-level fusion strategy that integrates the outcomes from multiple classifiers, including CNN. By managing the inherent information overlap and minimizing prediction errors, decision-level fusion enhances the reliability of our diagnostic tool. Fig. 7 illustrates the ensemble voting mechanism, detailing how classification results from all models are synthesized to improve diagnostic accuracy. Unlike data-level fusion, which can introduce significant noise, decision-level fusion combines the outcomes from independently trained models, thus harnessing a richer set of feature information while reducing the risk of accumulated prediction errors.

F. Data Analysis

In this study, we compiled a detailed dataset from the eye-movement data collected, focusing on key features identified as crucial for diagnosing dyslexia. As detailed in Table II, we analyzed 13 distinct features including Mean Fixation Duration (MFD), Standard Deviation of Fixation Duration (SDFD), and others critical for our analysis. For statistical comparison, we employed the Wilcoxon signed-rank test [41], suitable for small sample sizes and non-normally distributed data, which is typical in clinical studies. This test was used to assess whether the median difference between the dyslexic and non-dyslexic groups for each feature was zero. The P-values [42] reported in Table III indicate the probability of observing the data assuming the null hypothesis is true. Features with P-values less than 0.05 are considered statistically significant, suggesting that these features can reliably differentiate between normal and dyslexic individuals. This finding means that the differences observed in these features are statistically unlikely to occur by chance. The statistically significant features identified through the Wilcoxon signed-rank test ($P < 0.05$) contribute meaningfully to our analysis. These features were used as inputs in our decision-level fusion strategy, which combines outcomes from multiple classifiers, including CNN, to enhance the diagnostic accuracy of dyslexia identification.

In the latter part of our analysis, we focused on Case A and Case B to illustrate how our findings can be applied to individual participants. These cases represent two dyslexic individuals from our study group of 10 dyslexic and 4 non-dyslexic participants, highlighting the differences in reading patterns associated with varying degrees of dyslexia.

1) Case A, diagnosed with dyslexia, was classified as normal by the CNN classifier in the saliency map analysis. This individual exhibited occasional extended gaze durations, with a standard deviation of 1.73, just above one standard deviation from the norm, suggesting mild word formation difficulties. The other key metrics, such as mean and standard deviations of gaze time (0.09, 2.83, and 0.44, respectively), were all within one standard deviation of normal, indicating that Case A displayed typical reading patterns most of the time, with only a few deviations.
2) Case B, also diagnosed with dyslexia, demonstrated significantly longer gaze durations, with a mean of 0.16 and a standard deviation of 5.83, both exceeding two standard deviations from the norm. This suggests severe visual-perceptual impairments. Case B's re-gaze behavior and line-skipping tendencies were also much more pronounced compared to Case A. While some metrics, such as mean and standard deviation of gaze time (0.22 and 0.23, respectively), were closer to the norm, the overall reading behavior revealed more significant difficulties in visual perception and text navigation.

TABLE II
DEFINITION AND DESCRIPTION OF ALL FEATURES

Name	Description
MFD	Mean Fixation Duration
SDFD	Standard Deviation Fixation Duration
MS	Mean Saccade
SDS	Standard Deviation Saccade
CPM	Characters Per Minute
AFD	Average Fixation Duration
ASL	Average Saccadic Length
HLSD	High Frequency, Low Frequency Word Reading Speed Difference
GD	Gaze Duration
RG	Total Rereading Gaze
TRT	Total Reading Time
MRG	Mean Rereading Gaze
SDRG	Standard Deviation Rereading Gaze

TABLE III
P-VALUE OF ALL FEATURES

Feature Set/P-value	<0.05	>0.05	ALL
Eye Movement	4	9	13

Thus, Case A exhibited mild dyslexia, with gaze durations and reading patterns largely within normal limits, except for a few instances of extended gaze time, indicating minor word-formation difficulties and Case B displayed severe dyslexia, characterized by prolonged gaze durations, frequent re-gazing, and line-skipping, suggesting significant visual-perceptual difficulties. Through the analysis of these features and individual case studies, we demonstrate the ability of our approach to distinguish between varying degrees of dyslexia.

IV. RESULTS AND DISCUSSION

This section presents the performance of different machine learning classifiers applied to our dataset, along with the results from advanced models such as BERT and CNN. Additionally, we assess the effectiveness of a fusion model, which combines multiple classifiers using a voting mechanism to improve dyslexia diagnosis accuracy.

A. Performance of Machine Learning Classifiers on Eye Movement Features

We evaluated multiple classifiers in our machine learning analysis for dyslexia diagnosis, including Support Vector Machine (SVM), Random Forest (RF), K Nearest Neighbour

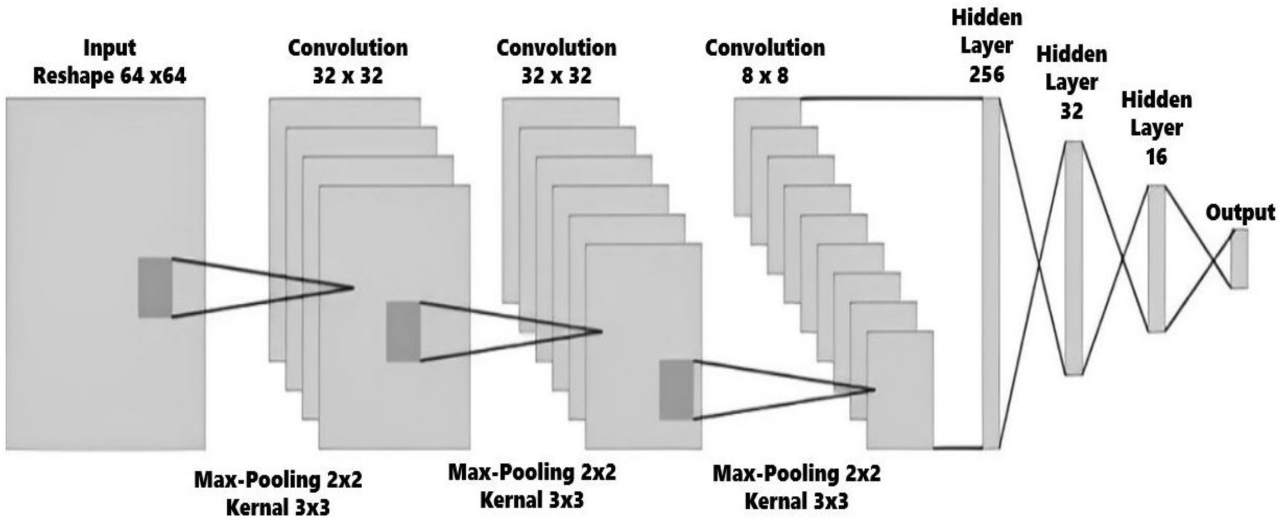

Fig. 6. CNN model detail schematic.

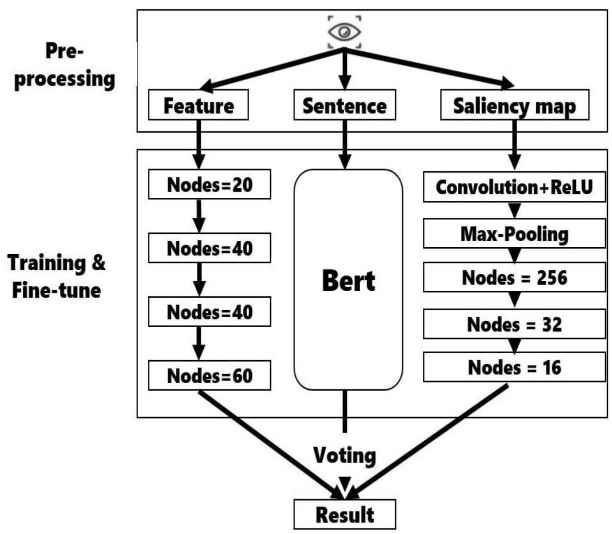

Fig. 7. Voting fusion model.

TABLE IV
THE CLASSIFICATION EFFECT FOR EACH ML-MODEL OF EYEMOVEMENT FEATURES

ML Model	Accuracy
SVM	0.91
Random Forest	1.0
KNN	0.76
Decision Tree	0.96
Gaussian NB	0.94
XGBoost	0.95
Deep Neural Network	0.91

(KNN), Decision Tree, Gaussian Naive Bayes, XGBoost, and Deep Neural Network (DNN). The performance of each classifier is summarized in Table IV.

The RF classifier achieved 100% accuracy under controlled conditions with a balanced dataset, indicating its strong ability to capture underlying relationships in the data, particularly when using key eye-tracking features like fixation duration, saccade amplitude, and fixation count. As shown in Table III, 4 out of 13 features exhibited statistically significant differences ($p < 0.05$), highlighting distinct eye movement behaviors that are critical in distinguishing dyslexic tendencies. These features underscore the utility of eye movement analysis in understanding dyslexic reading behaviors. In real-world settings with more diverse and unbalanced data, RF risks overfitting, especially with noisy features. To account for this, we applied 10-fold cross-validation and out-of-bag error estimation to ensure broader generalizability.

SVM and XGBoost also demonstrated high performance, with 91% and 95% accuracy, respectively. SVM required careful hyperparameter tuning to ensure optimal performance, while XGBoost faced challenges in handling complex data without L1/L2 regularization, which helped mitigate overfitting in the results. Both models provided robust insights into dyslexic reading patterns when dealing with clean, pre-processed data. Gaussian Naive Bayes reached 94% accuracy, performing particularly well in smaller datasets with continuous features. Despite its ability to work effectively under such conditions, Gaussian NB struggles with feature dependency, which reduces its effectiveness in more complex scenarios.

Meanwhile, Decision Tree exhibited 96% accuracy, reflecting its strength in handling both categorical and continuous data. However, Decision Tree models are prone to overfitting, which was mitigated by using cross-validation. KNN, however, achieved the lowest accuracy at 76%, as its sensitivity to noisy or irrelevant features made it less effective in this application. KNN often requires significant preprocessing to improve its accuracy in high-dimensional data. We observe that DNNs delivered 91% accuracy, excelling in capturing nonlinear patterns in the data. Despite its high complexity, DNN's performance improved significantly by reducing the number of hidden layers and nodes, which helped prevent overfitting.

B. Performance of BERT and CNN Models for Saliency Map and Cognitive Patterns

In addition to traditional classifiers, BERT was used to interpret dyslexic reading patterns by analyzing the grammatical structures and identifying areas of high cognitive load in

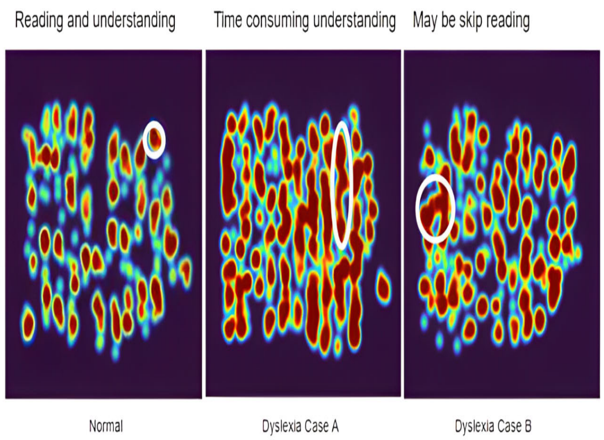

Fig. 8. Saliency maps of normal and two whole symptoms cases.

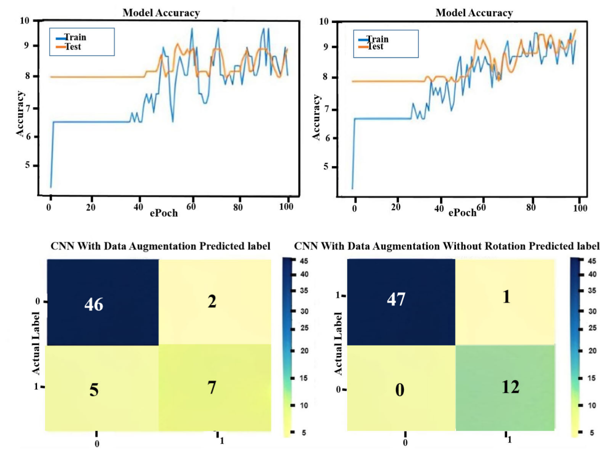

Fig. 9. Comparative analysis of CNN training with and without rotation augmentation.

readers. BERT's self-attention mechanism enabled it to focus on specific text regions where dyslexic individuals struggled the most.

For instance, Case A, which BERT initially misclassified, involved a dyslexic reader who had difficulty with word formation. While BERT accurately predicted several cases, in this particular scenario, BERT's prediction diverged from the visual data provided by CNN-generated saliency maps. This highlighted BERT's limitation when linguistic data alone is insufficient to capture the full scope of reading difficulties. Overall, BERT provided valuable insights into the linguistic complexities faced by dyslexic readers, but its performance was significantly enhanced when combined with CNN to account for both visual and cognitive factors.

CNNs were employed to analyze saliency maps generated from eye-tracking data, capturing the visual patterns of readers. Saliency maps indicated the areas where dyslexic readers spent more time, especially on specific words or phrases that posed difficulties, as illustrated by warm color peaks on the maps. Fig. 8 presents examples of saliency maps for normal readers and two dyslexic cases (Case A and Case B). Normal readers (left) exhibited typical reading behaviors with brief pauses to grasp content. Case A (middle) demonstrated prolonged fixations on certain areas, indicative of word formation difficulties, and spent more time on specific sections of text compared to the average reader. Case B (right) showed evidence of visual perceptual interference, potentially leading to skip reading or missing portions of text.

Our analysis revealed irregular reading patterns among dyslexic readers, such as frequent regressions and prolonged fixations, consistent with Rayner's findings [39] on eye movements in reading disorders. These patterns, distinct from non-dyslexic readers, highlight challenges in decoding and linguistic processing. Saliency maps provided a complementary perspective, illustrating heightened focus on specific words or phrases, corroborating Itti et al. [40] model of saliency-based visual attention. In some cases, BERT's predictions diverged from the outcomes indicated by the saliency maps. For example, BERT misclassified Case A, but the saliency map identified abnormalities related to word-formation difficulties, impacting attention patterns. Conversely, BERT correctly predicted Case B, but the saliency map did not reveal the pronounced visual perceptual interference observed in that case. These findings demonstrate that combining BERT's textual analysis with CNN's visual focus helps form a more robust and nuanced understanding of dyslexia. The individual performance of the BERT and CNN model in terms of the confusion matrix is shown in Fig. 10.

Figure 9 shows the impact of rotation augmentation on CNN performance. On the left, the confusion matrix reflects the performance of the CNN with rotation augmentation. The model achieved 46 correct classifications of non-dyslexic readers, with 2 false positives and 5 false negatives. While rotation augmentation was intended to simulate real-world variability in reading patterns, the results show that it slightly misled the model, making it less effective in certain cases where the visual attention was tilted or skewed by the augmented rotations.

On the right, the confusion matrix shows the results of the CNN without rotation augmentation. This model achieved 47 correct classifications of non-dyslexic readers and no false negatives, with only 1 false positive. Without rotation, the model performed better, demonstrating that tilting the saliency maps through rotation augmentation introduced a layer of complexity that hindered its ability to detect dyslexic patterns consistently.

In conclusion, rotation augmentation introduced unexpected challenges, slightly reducing the model's accuracy due to altered visual patterns, but it also enhanced the robustness in some scenarios by providing variability during training. The model without rotation augmentation performed better overall, as the visual features remained more aligned with typical reading behaviors, leading to fewer misclassifications.

C. Performance of Fusion Model (Voting Mechanism)

A fusion model was implemented to enhance diagnostic accuracy and minimize biases, combining the strengths of CNN, BERT, and DNN through a voting mechanism. This approach excels where individual models may be uncertain,

TABLE V
VOTING MODEL DETAILS

	Precision	Recall	F1-Score	Support
0(Dyslexia)	0.98	1.00	0.99	48
1(Normal)	1.00	1.00	1.00	12
Accuray			0.98	60
Macro Avg	0.96	0.99	0.97	60
Weighted Avg	0.98	0.98	0.98	60

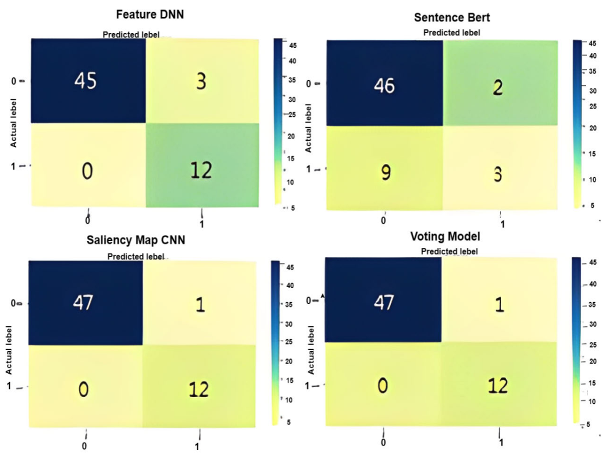

Fig. 10. Confusion matrices for DNN, BERT, CNN, and voting model classifications.

providing a robust consensus. Reducing false negatives is crucial in dyslexia diagnosis, as it ensures necessary educational interventions. One classifier may detect erratic eye movements, while another identifies subtle reading flow disruptions. Combined, the fusion model better identifies dyslexic cases that individual models might miss, reducing false negatives in our results. This approach is supported by [43], [44] on ensemble learning techniques, both demonstrating enhanced performance and reduced classification errors.

This ensemble approach achieved a consolidated accuracy of 98% as shown in Table V, balancing the strengths of each classifier to deliver more accurate predictions.By achieving higher precision, recall, and F1 scores, the fusion model demonstrates its ability to reduce both false positives and false negatives, which is crucial for ensuring reliable dyslexia diagnosis, particularly in real-world applications where minimizing diagnostic errors is critical.

In Fig. 10, the confusion matrix demonstrates the performance of the fusion model alongside individual models like DNN, BERT, and CNN. The confusion matrix shows that the DNN correctly classified 45 out of 48 non-dyslexic cases (labeled as 0) and all 12 dyslexic cases (labeled as 1). However, it misclassified 3 non-dyslexic cases as dyslexic, leading to a slight increase in false positives. The DNN performed well in handling the dyslexic group (zero false negatives), but it exhibited some bias towards over-predicting dyslexia in non-dyslexic participants.

The BERT model had more challenges than DNN, as evidenced by 9 false negatives—cases where dyslexic individuals were misclassified as non-dyslexic. BERT correctly classified 46 out of 48 non-dyslexic cases, but it struggled with dyslexic cases, correctly identifying only 3 out of 12, leading to 9 false negatives. This suggests that while BERT is capable of analyzing textual and cognitive patterns, it may not be as reliable when visual data (i.e., eye-tracking) is a more dominant indicator.

The CNN model achieved a near-perfect classification with just 1 false positive and no false negatives, correctly classifying 47 out of 48 non-dyslexic cases and all 12 dyslexic cases. This shows CNN's strength in recognizing visual patterns, such as prolonged fixations and regressions, which are key dyslexic indicators. As the confusion matrix for CNN and the fusion model are identical as shown in Fig. 8, it indicates that the CNN's predictions strongly influenced the final output of the voting mechanism.

The fusion model performed robustly in this context, but the results suggest that it relied heavily on the CNN model's output due to CNN's exceptional performance in detecting dyslexic patterns from eye-tracking data. The DNN model contributed by reducing false negatives, while BERT, despite being less reliable in identifying dyslexic cases, provided valuable insights into cognitive text processing. Ultimately, the voting mechanism ensured that the overall model balanced contributions effectively, although CNN's strength in visual data made it the dominant contributor. This analysis underscores the importance of combining both visual data (through CNN) and textual data (through BERT) in dyslexia diagnosis, with the fusion model acting as a safeguard to ensure comprehensive and accurate predictions.

V. CONCLUSION AND FUTURE WORK

This study presents a preliminary investigation into the use of VR and eye movement tracking for the diagnosis of dyslexia. Despite the small sample size, our findings demonstrate the potential of this innovative approach to enhance the accuracy and efficiency of dyslexia diagnosis. Our machine learning models, including DNN, BERT, and CNN, achieved high accuracy, with the fusion model achieving 98% accuracy. This approach significantly reduced false positives and negatives, especially with the fusion model, which integrated both visual and textual data. Saliency maps and eye movement analysis effectively highlighted reading anomalies specific to dyslexic individuals, such as prolonged fixations and frequent regressions. The integration of CNN and BERT provided a nuanced understanding of dyslexia, ensuring a more reliable diagnosis.

While the initial investment in VR equipment like the HTC Vive Pro with Tobii eye-tracking can be substantial, integrating training programs for educational staff will be crucial to ensure successful implementation and interpretation of results. Our methodology initially focused on Chinese scripts, is designed to be adaptable to other languages, aiming to ensure that the diagnostic tool is equitable and globally applicable. This sets the stage for enhancing the model's robustness and relevance across different linguistic and cultural contexts.

Looking forward, comprehensive clinical trials with larger and more diverse participant groups will be necessary to validate and refine our findings. Future research will extend this approach to diverse linguistic contexts and examine cultural differences in dyslexia manifestation to enhance the model's robustness and global relevance. Our future efforts will focus on addressing the identified limitations and expanding the applicability of our methodologies to achieve global educational equity in dyslexia diagnosis.

References

[1] NINDS Dyslexia Information Page. (2015). *National Institute of Neurological Disorders and Stroke National Institutes of Health*. Accessed: Jul. 1, 2022. [Online]. Available: https://www.ninds.nih.gov/health-information/disorders/learning-disabilities?search-term=%20Dyslexia

[2] J. M. Fletcher et al., "Classification of learning disabilities: An evidence-based evaluation," in *Identification of Learning Disabilities: Research to Practice*. Washington, DC, USA: Erlbaum Associates Publishers, 2002, pp. 185–250.

[3] L. Siegel, "Perspectives on dyslexia," *Paediatrics Child Health*, vol. 11, no. 9, pp. 581–587, 2006.

[4] R. Peterson and B. Pennington, "Developmental dyslexia," *Lancet*, vol. 379, no. 9830, pp. 1997–2007, 2012.

[5] J. J. S. Kooij, *Adult ADHD: Diagnostic Assessment and Treatment*, Springer, London, U.K., 2013, p. 83.

[6] (2018). *Statistics Department, Ministry of Health and Welfare Number of Persons With Disabilities By Category, County and City*. Accessed: Jul. 9, 2022. [Online]. Available: https://dep.mohw.gov.tw/dos/cp-5224-62359-113.html

[7] National Institutes of Health. (2022). *What Are Common Treatments for Reading Disorders*. Accessed: Jul. 5, 2022. [Online]. Available: https://www.nichd.nih.gov/health/topics/reading/conditioninfo/treatment

[8] R. Kunert and C. Scheepers, "Speed and accuracy of dyslexic versus typical word recognition: An eye-movement investigation," *Frontiers Psychol.*, vol. 5, p. 1129, Oct. 2014.

[9] J. A. G. Lum, M. T. Ullman, and G. Conti-Ramsden, "Procedural learning is impaired in dyslexia: Evidence from a meta-analysis of serial reaction time studies," *Res. Develop. Disabilities*, vol. 34, no. 10, pp. 3460–3476, Oct. 2013.

[10] (2018). *Instructions for the Identification of Students With Learning Disabilities*. Accessed: Jul. 9, 2022. [Online]. Available: http://teacher.mhups.tp.edu.tw/eweb/module/download/update/ew00000000104/file4840-14.pdf

[11] K. K. H. Chung, C. McBride-Chang, S. W. L. Wong, H. Cheung, T. B. Penney, and C. S.-H. Ho, "The role of visual and auditory temporal processing for Chinese children with developmental dyslexia," *Ann. Dyslexia*, vol. 58, no. 1, pp. 15–35, Jun. 2008.

[12] K. K. H. Chung and C. B. Lam, "Cognitive-linguistic skills underlying word reading and spelling difficulties in Chinese adolescents with dyslexia," *J. Learn. Disabilities*, vol. 53, no. 1, pp. 48–59, Jan. 2020.

[13] K. K. H. Chung, J. C. M. Lo, and C. McBride, "Cognitive-linguistic profiles of Chinese typical-functioning adolescent dyslexics and high-functioning dyslexics," *Ann. Dyslexia*, vol. 68, no. 3, pp. 229–250, Oct. 2018.

[14] H. Ranjbartabar, D. Richards, A. A. Bilgin, and C. Kutay, "First impressions count! The role of the Human's emotional state on rapport established with an empathic versus neutral virtual therapist," *IEEE Trans. Affect. Comput.*, vol. 12, no. 3, pp. 788–800, Jul. 2021.

[15] L. Bozgeyikli, A. Raij, S. Katkoori, and R. Alqasemi, "A survey on virtual reality for individuals with autism spectrum disorder: Design considerations," *IEEE Trans. Learn. Technol.*, vol. 11, no. 2, pp. 133–151, Apr. 2018.

[16] C. Meng and X. Zhao, "Webcam-based eye movement analysis using CNN," *IEEE Access*, vol. 5, pp. 19581–19587, 2017. [Online]. Available: https://ieeexplore.ieee.org/document/8046009

[17] Y. Mao, Y. He, L. Liu, and X. Chen, "Disease classification based on synthesis of multiple long short-term memory classifiers corresponding to eye movement features," *IEEE Access*, vol. 8, pp. 151624–151633, 2020.

[18] G. Rippon, "Trait and state EEG indices of information processing in developmental dyslexia," *Int. J. Psychophysiol.*, vol. 36, no. 3, pp. 251–265, Jun. 2000.

[19] H. Perera, M. F. Shiratuddin, and K. W. Wong, "Review of EEG-based pattern classification frameworks for dyslexia," *Brain Informat.*, vol. 5, no. 2, p. 4, Dec. 2018.

[20] O. L. Usman, R. C. Muniyandi, K. Omar, and M. Mohamad, "Advance machine learning methods for dyslexia biomarker detection: A review of implementation details and challenges," *IEEE Access*, vol. 9, pp. 36879–36897, 2021.

[21] A. Vaswani et al., "Attention is all you need," 2017, *arXiv:1706.03762*.

[22] C. A. N. Knoop-van Campen, E. Segers, and L. Verhoeven, "Effects of audio support on multimedia learning processes and outcomes in students with dyslexia," *Comput. Educ.*, vol. 150, Jun. 2020, Art. no. 103858.

[23] Y. Lin, H. Ma, Z. Pan, and R. Wang, "Depression detection by combining eye movement with image semantics," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 269–273.

[24] S. Choudhary, K. Tandon, R. Agarwal, and N. Chatterjee, "MTL782_IITD at CMCL 2021 shared task: Prediction of eye-tracking features using BERT embeddings and linguistic features," in *Proc. Workshop Cognit. Modeling Comput. Linguistics*, 2021, pp. 114–119.

[25] A. Elnakib, A. Soliman, M. Nitzken, M. F. Casanova, G. Gimel'farb, and A. El-Baz, "Magnetic resonance imaging findings for dyslexia: A review," *J. Biomed. Nanotechnol.*, vol. 10, no. 10, pp. 2778–2805, Oct. 2014.

[26] U. Kuhl et al., "The emergence of dyslexia in the developing brain," *NeuroImage*, vol. 211, May 2020, Art. no. 116633.

[27] E. Paulesu, L. Danelli, and M. Berlingeri, "Reading the dyslexic brain: Multiple dysfunctional routes revealed by a new meta-analysis of PET and fMRI activation studies," *Frontiers Hum. Neurosci.*, vol. 8, p. 830, Nov. 2014.

[28] K. Rayner, A. Pollatsek, J. Ashby, and C. Clifton Jr., *Psychology of Reading*. London, U.K.: Psychology Press, 2012.

[29] A. Jothi Prabha and R. Bhargavi, "Predictive model for dyslexia from fixations and saccadic eye movement events," *Comput. Methods Programs Biomed.*, vol. 195, Oct. 2020, Art. no. 105538.

[30] M. De Luca, E. Di Pace, A. Judica, D. Spinelli, and P. Zoccolotti, "Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia," *Neuropsychologia*, vol. 37, no. 12, pp. 1407–1420, Nov. 1999.

[31] K. I. Schattka, R. Radach, and W. Huber, "Eye movement correlates of acquired central dyslexia," *Neuropsychologia*, vol. 48, no. 10, pp. 2959–2973, Aug. 2010.

[32] P. Raatikainen, J. Hautala, O. Loberg, T. Kärkkäinen, P. Leppänen, and P. Nieminen, "Detection of developmental dyslexia with machine learning using eye movement data," *Array*, vol. 12, Dec. 2021, Art. no. 100087.

[33] J. A. Prabha, R. Bhargavi, and B. Harish, "Predictive model for dyslexia from eye fixation events," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1s3, pp. 235–240, Dec. 2019.

[34] C. I. Eke, A. A. Norman, and L. Shuib, "Context-based feature technique for sarcasm identification in benchmark datasets using deep learning and BERT model," *IEEE Access*, vol. 9, pp. 48501–48518, 2021.

[35] R. Cai et al., "Sentiment analysis about investors and consumers in energy market based on BERT-BiLSTM," *IEEE Access*, vol. 8, pp. 171408–171415, 2020.

[36] J. He and H. Hu, "MF-BERT: Multimodal fusion in pre-trained BERT for sentiment analysis," *IEEE Signal Process. Lett.*, vol. 29, pp. 454–458, 2022.

[37] S. Lee, D. K. Han, and H. Ko, "Multimodal emotion recognition fusion analysis adapting BERT with heterogeneous feature unification," *IEEE Access*, vol. 9, pp. 94557–94572, 2021.

[38] A. Frid and Z. Breznitz, "An SVM based algorithm for analysis and discrimination of dyslexic readers from regular readers using ERPs," in *Proc. IEEE 27th Conv. Electr. Electron. Eng. Isr.*, Nov. 2012, pp. 1–4.

[39] K. Rayner, "Eye movements in reading and information processing: 20 years of research," *Psychol. Bull.*, vol. 124, no. 3, pp. 372–422, 1998.

[40] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[41] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945.

[42] J. Miller, "The role of p-values in hypothesis testing: A review," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 7, pp. 1879–1888, Jan. 2019.

[43] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Mach. Learn.*, vol. 51, no. 2, pp. 181–207, 2003.

[44] R. Hao, K. Namdar, L. Liu, M. A. Haider, and F. Khalvati, "A comprehensive study of data augmentation strategies for prostate cancer detection in diffusion-weighted MRI using convolutional neural networks," *J. Digit. Imag.*, vol. 34, no. 4, pp. 862–876, 2021.