Efficient Dyslexia Detection Using a Subset of Facial Features: a Comparative Analysis with ANN and CNN

Günet Eroğlu and Mhd Raja Abou Harb

August 26, 2024

Efficient Dyslexia Detection Using a Subset of Facial Features: A Comparative Analysis with ANN and CNN

Günet Eroğlu1[0000-0001-8382-8417] and MHD Raja Harb2[0000-0002-8637-5473]

[1] Bahçeşehir University, Istanbul, Turkey
[2] Işık University, Istanbul, Turkey
gunet.eroglu@healthmobilesoftware.com

Abstract. Dyslexia, a learning disability that affects reading and writing skills, has been the subject of extensive research, particularly in the field of machine learning-based classification. In this study, we focused on a subset of features extracted using Google ML Kit, to enhance the efficiency of dyslexia classification. The selected features resulted in about 97% of the total number of features that had been collected. We employ two classification methods, Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN), to evaluate the effectiveness of this feature subset. Preliminary results indicate that the selected features contribute significantly to the classification accuracy, with the CNN model outperforming the ANN in most scenarios. This focused approach not only streamlines the feature selection process but also demonstrates the potential for more targeted and efficient dyslexia detection methods. Our findings suggest that reducing the feature set while maintaining high classification performance is feasible, paving the way for more practical applications in real-world settings.

Keywords: Dyslexia Classification, Learning Disabilities, Supervised Machine Learning, Feature Subset Selection, Google ML Kit.

1 Introduction

Word identification, spelling, and decoding skills are frequently affected by dyslexia, a unique learning disorder that predominantly impairs reading and writing abilities. People who have dyslexia may have trouble with phonological processing, which is necessary for associating sounds with letters and words, even in the presence of normal intelligence and sufficient educational opportunity. This neurological disorder is associated with variations in the way the brain interprets language rather than visual or auditory abnormalities. Since dyslexia affects 5–10% of the population, early detection and treatments are essential for both academic achievement and general development, according to research. A deeper understanding of the neurological and genetic roots of dyslexia has been made possible by developments in neuroimaging and genetic research, underscoring the significance of individualized learning approaches and tailored instructional strategies [1-3].

Early identification of dyslexia is crucial for implementing appropriate interventions and support strategies. However, detecting dyslexia can be challenging, as traditional methods rely on subjective evaluations and standardized tests that are time-consuming and may not capture the full range of dyslexia symptoms. Moreover, these methods may be influenced by cultural and linguistic factors, leading to potential biases in diagnosis [4].

Classical methods of detecting dyslexia typically involve a combination of standardized tests, observational assessments, and interviews conducted by educational psychologists or specialists. These approaches often assess various aspects of reading and writing, including phonological awareness, decoding skills, reading fluency, and comprehension. Common tools include the Dyslexia Screening Test (DST) and the Wechsler Individual Achievement Test (WIAT), which are used to identify patterns consistent with dyslexia. However, these methods are often time-consuming, requiring multiple sessions to gather comprehensive data, and involve significant effort from both the assessor and the individual being tested. Moreover, these assessments are usually performed in clinical or educational settings, which can add to the logistical challenges, especially for large-scale screening. The labor-intensive nature of these traditional methods underscores the need for more efficient, automated approaches that can streamline the detection process while maintaining accuracy [1, 5-6].

With the rapid advancement in the field of artificial intelligence, the feasibility of detecting dyslexia through supervised machine-learning techniques has gained significant attention. These methods offer the potential for more accessible and cost-effective solutions to this complex classification problem. Various types of features have been explored for dyslexia detection, including electroencephalography (EEG) data, functional magnetic resonance imaging (fMRI) scans, and eye movement patterns. EEG data, for instance, captures brain wave activity that has been shown to differ between individuals with and without dyslexia, providing a valuable signal for early detection [7]. fMRI has been used to identify differences in brain activity during reading tasks, offering insights into the neural mechanisms underlying dyslexia [8]. Eye movement patterns, such as fixation duration and saccadic behavior, have also been studied, as individuals with dyslexia often exhibit distinctive reading behaviors that can be detected through eye-tracking technology [9]. However, the collection of these features typically requires expert supervision and specialized equipment, making the process resource-intensive and less scalable for widespread screening applications.

The current study aims to enhance dyslexia detection by focusing on a carefully selected subset of these facial features. Additionally, we shift our emphasis toward deep neural networks, specifically exploring the performance of Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) to achieve more accurate and efficient dyslexia classification.

This paper is organized as follows: Section 2 reviews related work in the field. Section 3 provides background information and outlines the methodologies used in this study. Section 4 presents the results of our experiments, while Section 5 discusses their implications. Finally, Section 6 concludes the paper and suggests directions for future research.

2 Related Works

To focus our exploration of dyslexia detection, we have chosen to concentrate on eye movements, which have consistently proven to be a reliable indicator of reading difficulties. Eye movement analysis offers a non-invasive and practical method for identifying dyslexic patterns. In a 2023 study, Sekhar and Chandrashekar employed XGBoost, Support Vector Machine (SVM), and Random Forest (RF) classifiers to detect dyslexia in 185 students using eye movement data collected while reading a passage. The data collection was conducted using the Ober-2 eye-tracking goggles, which captured detailed information on eye movement positions, saccades, and fixations. Additionally, they analyzed the duration of the reading process, the average and standard deviation of eye movement positions, the distance between eye positions, and the maximum value between any two positions (Left-eye to Right-eye). The XGBoost classifier achieved the highest accuracy, with a score of 95% [10].

In their 2023 study, Vajs et al. aimed to bridge the gap between different study designs by developing a machine learning-based pipeline evaluated on two distinct eye-tracking datasets—training on one and testing on the other, and vice versa. The first dataset comprised 30 participants (15 dyslexic and 15 control) aged 7 to 13, with a gender distribution of 19 females and 11 males. The second dataset included 185 participants (97 dyslexic and 88 control) aged 9 to 10, with 145 males and 40 females. Data for the first dataset was collected using the SMI RED-m 120 Hz portable remote eye-tracker, while the Ober-2™ eye-tracking tool was employed for the second dataset. The reading texts for the first dataset were in Serbian, whereas Swedish texts were used for the second dataset. Both datasets were converted into grayscale images by applying various time window configurations to parse the signals and plot the data on a 2D plane. These images were then used to train an Autoencoder neural network, with the reconstruction error serving as features to describe each instance in the training and testing sets. Various machine learning algorithms were trained on the extracted features, and the models were evaluated on the testing feature dataset. The study achieved classification accuracies of 85.6% when testing on Serbian readers' data and 82.9% on Swedish readers' data using Logistic Regression (LR) [11].

In their 2023 study, Shalileh et al. focused on identifying dyslexia in school pupils by leveraging eye movement and demographic data through artificial intelligence. The research utilized a comprehensive dataset that combined eye-tracking metrics, such as fixation duration, saccade amplitude, and pupil size, with demographic information, including age, gender, and reading proficiency levels. The total number of participants was 307, and the Russian language was used in the experiment. The study employed a variety of machine learning algorithms to analyze the data, ultimately aiming to distinguish between dyslexic and non-dyslexic students. By integrating these diverse data sources, the researchers sought to improve the accuracy of dyslexia detection. The results demonstrated the effectiveness of using a multifaceted approach, with the models achieving significant classification accuracy, highlighting the potential of artificial intelligence in supporting early dyslexia identification in educational settings. The highest accuracy reached was 93.4% and it was scored by Multi-Layer Perceptron (MLP) [12].

The review of previous research indicates that there is possible opportunity in improving dyslexia detection through machine learning techniques. In this study, we aim to address enhancing the efficiency of the implementation process. We will focus on a subset of the features utilized in the data collected through Google ML Kit, which we believe will streamline the process and reduce computational demands. Additionally, by employing the Google ML Kit, our approach promises to be more cost-effective, as it eliminates the need for specialized equipment to collect data, making dyslexia detection more accessible and practical.

3 Background and Methodology

3.1 The “Eye Zenith” mobile app

The "Eye Zenith" app is an Android mobile application designed for Turkish-speaking users, utilizing the Google ML Kit API [13]. The app allows children to read one of three short Turkish stories, each taking approximately one minute. During this reading period, the app uses a Samsung Galaxy S20 FE to capture facial landmarks through its 32-megapixel front camera, ensuring high-resolution images for precise identification of facial features. For each story, the app records 1,000 facial landmark data points. This data is then collected and stored in MongoDB, an open-source, cross-platform document-oriented NoSQL database that operates via a web service. Each child participates in the experiment by reading three age-appropriate Turkish stories. Fig. 1 illustrates the app’s user interface during the reading process, with the page shown translated into English for better understanding.

3.2 Dataset Details

The total number of the used features in the training model is 288 which can be summarized in Table 1. In the preprocessing phase, zero values of the features have been replaced with the per-class median value for the features. A standard scaler had also been used in the preprocessing phase. The total number of records in the non-dyslexic cases was 22003 while 49339 records were scored related to the dyslexic cases. Since the number of records does not reflect the universal cases of dyslexia, it has been changed to the ratio of 90:10 for non-dyslexic and dyslexic cases respectively. Random-Over-Sampler was used to change the number of records. After prepossessing steps, the dataset was randomly split into training and testing with the ratio 80:20 respectively.

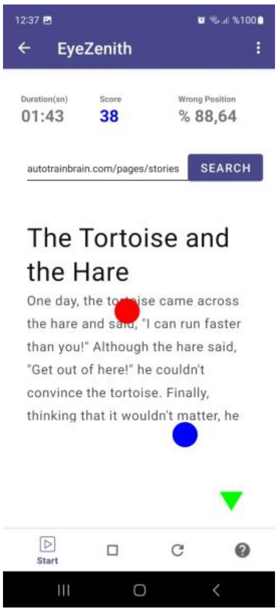

Fig. 1. Eye Zenith app user interface.

Table 1. Features overview.

Features group	Description	Number of features
Face detection	These features locate the face detected.	3 features contain the top, left, and bottom of the face.
Face Orientation	It contains the angle of the detected face.	3 features contain Euler x, Euler y, and Euler z.
Landmarks	The position of 10 main landmarks of the face.	20 features each 2 features containing the X and Y coordinates of the landmark.
Classifications	They contain some classification results for opening the eyes and smiling probabilities	3 features contain the probability of opening the right and left eyes and the probability of smiling.
Contours	It detects the contours of the face which were used to detect the landmarks	268 features every 2 features contain the X and Y coordinates of the contour.

3.3 Feature Selection

To enhance the implementation time of the proposed classifiers, a subset of features from Table 1 was utilized. Specifically, features related to face detection, face orientation, and classifications were selected, resulting in a 97% reduction in the total feature set. This subset was chosen because these features effectively represent the full range of data, providing a comprehensive overview while significantly reducing computational demands.

3.4 Classification Models

Because of the great performance that ANN classifiers achieved in similar research problems, we are going to use this classifier it in our experiment. A feedforward neural network with dropout regularization using Keras was used. The model features an input layer with 32 units and ReLU activation, followed by L2 regularization to prevent overfitting, and a dropout layer with a 25% rate to further reduce overfitting. A second dense layer with 16 units and ReLU activation is similarly regularized and followed by another dropout layer. The final layer is a single unit with sigmoid activation for binary classification. The model is compiled with the Adam optimizer and binary cross-entropy loss, and trained for 25 epochs with a batch size of 32, incorporating a 20% validation split. After training, the model is evaluated on the test set to obtain loss and accuracy metrics. The execution time for the training process is measured by recording the start time before training and calculating the elapsed time afterward.

Moreover, the CNN model is designed for sequence data. It starts with a convolutional layer that applies a series of filters to the input data, extracting key features and using ReLU activation to introduce non-linearity. This is followed by a max-pooling layer that reduces the dimensionality of the data while retaining important features. The output from the convolutional and pooling layers is then flattened into a single vector, which is fed into fully connected dense layers. The dense layers include one with a larger number of units and ReLU activation to capture complex patterns and a final layer with a single unit and sigmoid activation for binary classification. The model is compiled with the Adam optimizer and binary cross-entropy loss function and is trained over a specified number of epochs with a certain batch size.

Table 2 shows the evaluation measurements that are used to evaluate the models and a brief description for each one, knowing that TP is true positive which represents the number of dyslexic cases that were correctly predicted, FP is false positive, which represents the number of predicted cases of dyslexia, but they are not, FN is false negative which represents the number of predicted cases as non-dyslexic, but they are not, and TN is true negative which represents the number of correctly predicted cases as non-dyslexic.

One of the major charts that are used for evaluating machine learning models is the Receiver Operating Characteristic (ROC) curve. For any possible test or combination of tests, ROC curves are used to graphically show the connection/trade-off between sensitivity and specificity. In addition, the area under the ROC curve (AUC) indicates the benefit of using the test(s) in question. According to Zheng and Alice (2015), AUC shows how many correct positive classifications can be gained by allowing for more and more false positives. The higher the AUC value the better the machine learning model is in the studied problem.

Table2. Evaluation measurements.

Measurement	Description	Formula
Accuracy	It is determined by dividing the total number of correct predictions by the total number of predictions.	$Acc = \frac{TP + TN}{total\ samples}$
True Positive Rate (Sensitivity)	It displays the percentage of positive data points that were correctly predicted to be positive.	$Sensitivity = \frac{TP}{FN + TP}$
True Negative Rate (Specificity)	It displays the percentage of negative data points that were correctly predicted to be negative.	$Specificity = \frac{TN}{FP + TN}$
Precision	It is calculated by dividing the number of correct positive results by the number of positive results predicted by the classifier.	$Precision = \frac{TP}{TP + FP}$
False Positive Rate (FPR)	It displays the proportion of negative data points considered positive in the prediction.	$False\ Positive\ Rate = \frac{FP}{TN + FP}$

4 Results

We utilized Google Colab, a cloud-based development environment that offers a free and interactive platform for running Python code within a Jupyter Notebook format. This platform allows users to write, execute, and share Python code directly in a web browser, eliminating the need for local installation or setup. Powered by Google's cloud infrastructure, Google Colab provides high-performance computing resources, including CPUs, GPUs, and TPUs, enabling efficient execution of code at scale. All results presented in this section have been rounded to four decimal places.

Table 3 shows the machine learning results, while Fig. 2 and Fig. 3 show the ROC curves and AUC for ANN and CNN classifiers respectively.

Table 4 compares the reached results with the discussed previous works.

Table3. Machine learning models results.

ML model	Time (seconds)	Accuracy	Sensitivity	Specificity	Precision	FPR
ANN	94.1912	0.9489	0.9739	0.9262	0.9227	0.0738
CNN	147.8721	**0.9646**	0.9737	0.9557	0.9551	0.0443

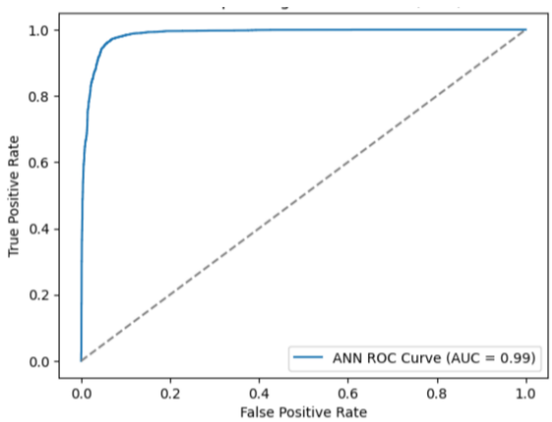

Fig. 2. ROC curve and AUC value of ANN classifier.

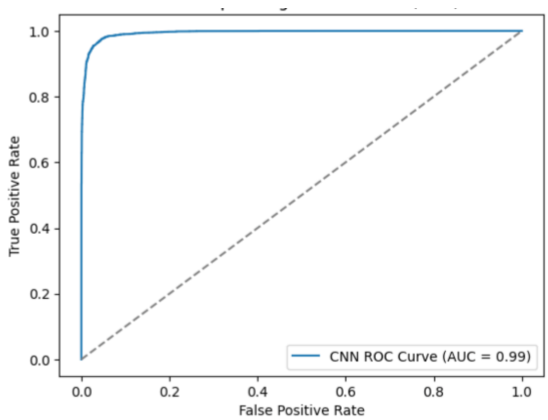

Fig. 3. ROC curve and AUC value of CNN classifier.

Table3. Comparing the results with the results of the previous works.

Used approach	reference	Highest reached accuracy	The machine learning model that reached this score

Eye movements feature	[11]	95%	The XGBoost classifier
	[12]	85.6%	The LR classifier
	[13]	93.4%	MLP
Face features	Our approach	96.46%	CNN classifier

5 Discussion

In this study, we addressed the problem of dyslexia detection using supervised machine learning models, focusing on improving the performance by using subset of the collected dataset. The approach involved utilizing a subset of the original feature set, leading to a significant 97% reduction in the number of features. Despite this reduction, the results demonstrate strong performance, with the ANN classifier achieving 94.89% accuracy and the CNN classifier reaching 96.46% accuracy. The implementation time was also notably efficient, requiring only 1.5 minutes for the ANN and 2.5 minutes for the CNN. Notably, the ROC curves for both models, as shown in Figures 2 and 3, exhibited no significant differences, with both models achieving similar AUC values.

When comparing our proposed solution to previous works that utilized eye movement features, as outlined in Table 3, our approach outperforms these earlier methods, highlighting the effectiveness of the reduced feature set.

Despite the promising results achieved in this study, there are several limitations that should be acknowledged. First, while the reduction in the feature set significantly improved implementation time, it may have also led to a slight decrease in accuracy. This trade-off between efficiency and accuracy suggests that further optimization may be necessary to maintain high classification performance while minimizing computational costs. Additionally, the generalizability of the proposed models may be limited by the specific dataset used, as the feature set was tailored to a particular set of tasks and may not capture all relevant aspects of dyslexia detection across diverse populations. Moreover, the reliance on Google ML Kit for data collection, although cost-effective, may introduce variability in the quality of the data, as it depends on the hardware and environmental conditions during data capture. Future work should explore the robustness of the models across different datasets and investigate ways to enhance feature selection to balance accuracy and efficiency more effectively.

6 Conclusion

In this study, we addressed the challenge of dyslexia detection by employing supervised machine learning models. By utilizing a significantly reduced feature set—achieving a 97% reduction—we were able to maintain high accuracy levels, with the ANN classifier reaching 94.89% and the CNN classifier achieving 96.46%. The findings suggest that feature reduction can be an effective strategy in enhancing the operational efficiency of machine learning models in dyslexia detection without severely compromising accuracy. Future research should continue to refine feature selection techniques and explore the applicability of the proposed approach across more diverse datasets and settings.

References

1. Shaywitz, S. E., Shaywitz, J.: Overcoming dyslexia: Completely revised and updated. Hachette UK (2020).
2. Peterson, R. L., Pennington, B. F.: Developmental dyslexia. The lancet, 379(9830), pp. 1997-2007 (2012).
3. Snowling, M. J., Hulme, C.: Annual Research Review: The nature and classification of reading disorders–a commentary on proposals for DSM-5. Journal of child psychology and psychiatry, 53(5), pp. 593-607 (2012).
4. Snowling, M. J., Hulme, C., Nation, K.: Defining and understanding dyslexia: past, present and future. Oxford Review of Education, 46(4), pp. 501-513 (2020).
5. Catts, H. W., Nielsen, D. C., Bridges, M. S., Liu, Y. S., Bontempo, D. E.: Early identification of reading disabilities within an RTI framework. Journal of learning disabilities, 48(3), pp. 281-297 (2015).
6. Elliott, J. G., Grigorenko, E. L.: The dyslexia debate (No. 14). Cambridge University Press (2014).
7. Lehongre, K., Ramus, F., Villiermet, N., Schwartz, D., Giraud, A. L.: Altered low-gamma sampling in auditory cortex accounts for the three main facets of dyslexia. Neuron, 72(6), pp. 1080-1090 (2011).
8. Pugh, K. R., Mencl, W. E., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Constable, R. T., Gore, J. C.: The angular gyrus in developmental dyslexia: task-specific differences in functional connectivity within posterior cortex. Psychological science, 11(1), pp. 51-56 (2000).
9. Hawelka, S., Gagl, B., Wimmer, H.: A dual-route perspective on eye movements of dyslexic readers. Cognition, 115(3), pp. 367-379 (2010).
10. Sekhar, S. R., Chandrashekar, S.: Eye Movement Interpretation for Detecting Dyslexia Using Machine Learning Techniques. International Journal of Computing and Digital Systems (2023).
11. Vajs, I. A., Kvaščev, G. S., Papić, T. M., Janković, M. M.: Eye-tracking image encoding: Autoencoders for the crossing of language boundaries in developmental dyslexia detection. IEEE Access, 11, pp. 3024-3033 (2023).
12. Shalileh, S., Ignatov, D., Lopukhina, A., & Dragoy, O.: Identifying dyslexia in school pupils from eye movement and demographic data using artificial intelligence. Plos one, 18(11), (2023).
13. Auto Train Brain Homepage: https://www.autotrainbrain.com), last accessed 11/08/2024.