Dyslexia detection in children using eye tracking data based on VGG16 network

Ivan Vajs
University of Belgrade – School of Electrical Engineering and Innovation center, School of Electrical Engineering in Belgrade
Belgrade, Serbia
ivan.vajs@ic.etf.bg.ac.rs

Vanja Ković
Laboratory for Neurocognition and Applied Cognition, Faculty of Philosophy, University of Belgrade
Belgrade, Serbia
vanja.kovic@f.bg.ac.rs

Tamara Papić
University Singidunum,
Belgrade, Serbia
tpapic@singidunum.ac.rs

Andrej M. Savić
University of Belgrade – School of Electrical Engineering
Belgrade, Serbia
andrej_savic@etf.bg.ac.rs

Milica M. Janković
University of Belgrade – School of Electrical Engineering
Belgrade, Serbia
piperski@etf.bg.ac.rs

***Abstract*—Considering the negative impact dyslexia has on school achievements, dyslexia diagnosis and treatment are found to be of great importance. In this paper, a deep convolutional neural network was developed to detect dyslexia in children ages 7-13, based on gathered eye tracking data. The children read a text written in Serbian on 13 different color configurations (including background and overlay color variations) and the raw gaze coordinates gathered during the trials were formatted into colored images and used to train a deep learning model based on the VGG16 architecture. Several configurations of the convolutional neural network were evaluated, as well as several trial segmentation configurations in order to provide the best overall result. The method was evaluated using subject-wise cross-validation and an accuracy of 87% was achieved. The obtained results show that a combination of convolutional neural network and visual encoding of the eye tracking data shows promising results in dyslexia detection with minimal preprocessing.**

Keywords— dyslexia, reading, eye tracking, convolutional neural networks, classification

I. Introduction

Up to 20% of the general population are found to exhibit some degree of reading difficulties [1], while about 7% of people are affected heavily enough to qualify for a dyslexia diagnosis [2]. Early diagnostics is more than important for the benefit of the child, but also for the benefit of broader society, given that dyslexia is linked to an increased risk of low school achievement and even school dropout, and other adverse outcomes. It is important to note that there is a particular value of developing an objective diagnostic tool for dyslexia in the shallow orthographic systems (with high correspondence between phonemes and letters). In those cases, diagnosing dyslexia based on performance on behavioral reading tests can be even more challenging, and thus employing measures which can tap into the process, and not only the outcomes of reading become even more advantageous.

Different methods for dyslexia detection have been implemented in recent studies. The analysis of eye tracking during reading have been of particular interest, as they provide a direct insight into the visual attention path of the subject, which can be analyzed to detect dyslexic tendencies [3], [4].

In [5] the authors claim to have performed the first attempt at dyslexia classification based on eye tracking features. The dataset in the paper contained data from 97 subjects, each reading 12 different texts in Spanish, each text with a different font type. All of the texts were presented on a white background with black letters. Each was represented by 12 features, including the age of the participant, number of visits to the region of interest, mean duration of fixations, total number of fixations, etc. A Support Vector Machine (SVM) classification model was trained and evaluated using 10-fold cross-validation and the overall accuracy of 80.18% was achieved.

The authors in [6] have performed dyslexia classification based on eye tracking features as well. They included a larger number of eye tracking features and performed the analysis on a dataset containing data from 185 subjects, where each subject read 8 lines of text (in Swedish) from a white piece of paper with high contrast. A total of 168 eye tracking features were considered, including data regarding the progressive and regressive fixations and saccades, i.e. the duration of an event, the distance spanning the event, the maximum range between any two positions, the accumulated distance over all subsequent positions, etc. Due to the large number of features, a feature selection algorithm was implemented and an SVM model was trained for dyslexia detection. The model was evaluated using 10-fold stratified cross-validation, repeated 100 times in order to stabilize the estimates across the random partitions of the dataset. The best obtained accuracy was 95.6% showing that a large pool of considered features can ensure an efficient dyslexia detection.

The analysis of eye tracking data for the purpose of dyslexia detection was taken a step further in [7] using a holistic approach. The dataset used in the paper was the same as the one used in [6] but rather than extracting features, the paper used raw measurements, i.e. gaze coordinates. The classification was performed using a convolutional neural network and a variation of the cross-validation method in order to evaluate the model. A high accuracy of 96.6% was achieved, thus proving that a convolutional neural network can be successfully implemented in order to classify dyslexic subjects based on raw eye tracking data.

Considering the research that has been performed in the field of eye tracking and dyslexia, it can be concluded that the

This research was supported by the Ministry of Education, Science and Technology Development of Serbia, Belgrade, Serbia.

features describing eye tracking events, or even raw eye tracking data can be successfully used to detect dyslexic tendencies in subjects. Considering this, the goal of this paper was to introduce a new approach on detecting dyslexia using raw eye tracking measurements on a dataset containing data from subjects reading text written in Serbian on different color configurations. The proposed algorithm is based on visualization of reading data in the form of colored images and uses a 2D convolutional neural network (based on the VGG16 architecture) that have been proven to be effective in image processing in a wide range of areas [8]–[11].

II. Method

A. Dataset

The dataset analyzed in this paper contains recordings of 30 subjects (19 female, 11 male, ages 7-13), of which 15 subjects were diagnosed with dyslexia and 15 participants were control subjects. The dataset and the detailed experimental protocol was described in [12]. However, six subjects from the original dataset from [12] were not adequate for this type of analysis due to a lower number of data segments and a short reading time (less than 5 s) indicating insufficient focus on the displayed text, resulting in a total number of 30 subjects that are used in this paper. During the experiment, the subjects were instructed to read the 13 segments of the text (from the standardized story for elementary school "St Sava and the villager without happiness") displayed on the computer screen, and to press the space bar when they have finished reading of each segment to start the display of the next segment of the text. The text segments had different color configurations in the sense that each segment had either black text with a colored background or a transparent colored overlay over a white background with black letters. The color order was randomly selected. The 6 colors used were: red, orange, yellow, turquoise, blue and purple, each of them used both as a background (B) and an overlay (O), giving 12 configurations total with the white color background configuration being the 13th. The sensor hub was used for multimodal monitoring of different biometric parameters of the subject: heartrate, electroencephalography (EEG), electrodermal activity (EDA) and eye tracking data, with the goal of observing the cognitive, behavioral and emotional response of subjects when reading on different color configurations However, this study focused only on the eye tracking data recorded by SMI RED-m 120 Hz portable remote eye tracker (iMotions, Copenhagen, Denmark). The observed data was stored in the form of a series of gaze coordinate pairs (*x,y*) per text segment (so called "trial").

B. Algorithm description

The overall methodology for eye tracking data preprocessing and classification is given in Fig. 1, with each step described in more detail in the following paragraphs.

1) Data preprocessing and visualization

Data preprocessing involved the elimination of eye blinks and points where the gaze was not registered, based on the status signal obtained directly from the eye tracker, as the first step. Secondly, each trial of each subject was divided into n nonoverlapped data instances of equal length. Three different values for n were considered in this paper (3, 5 and 10) in order to investigate its impact on the classification performance. Each of the data instances obtained through the trial division was visualized in a two-dimensional plane showing the trace of the subject's gaze in that particular time window. The data instance of the gaze was plotted using color coding, more precisely the jet color map, so that each line connecting two subsequent points (x_{k-1}, y_{k-1}) and (x_k, y_k) is colored in accordance to its length, Fig. 2. (longer distance between the points representing fast movements and shorter distance between the points representing slower ones).

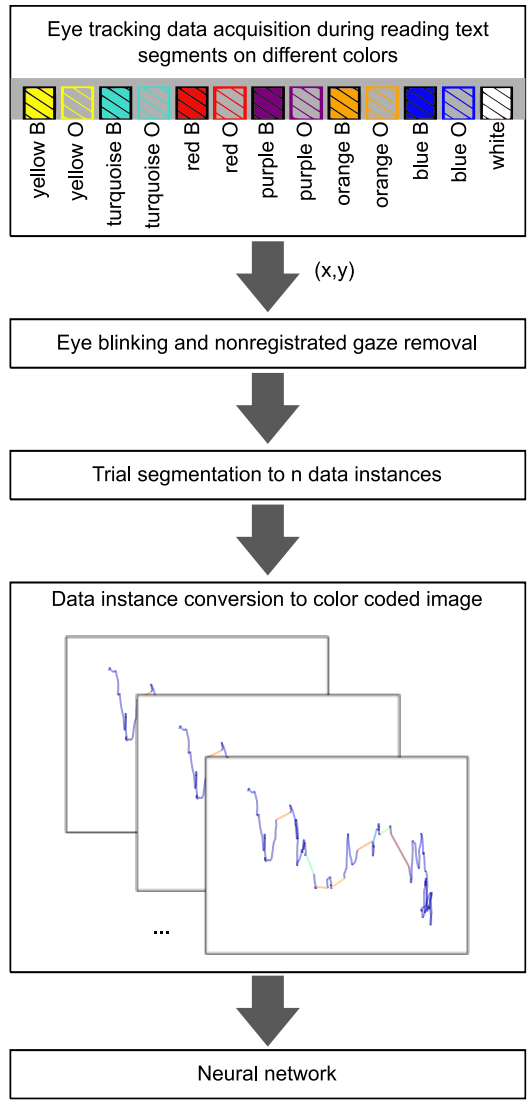

Fig. 1. The block diagram of the suggested algorithm.

This color-coded format makes it easier to differentiate quick and slow eye movements as the sample rate is fixed. The segment's color, therefore, corresponds to the movement velocity. The implemented color coding of each of the data sequences can also serve as a substitute for eye tracking event detection, since longer/faster movements will be displayed in a different color than the short ones. Each plotted image was used as a single data instance, in combination with the appropriate label (dyslexic or control) to form the input dataset which was used for the training and evaluation of the deep learning models.

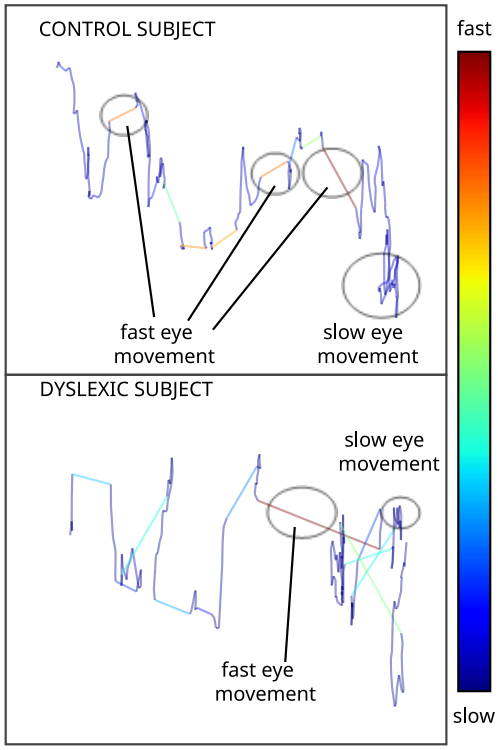

Fig. 2. Data visualization of data instance for control and dyslexic subject.

2) Deep learning algorithm

Several configurations of a deep convolutional neural network were implemented with the goal of obtaining the best possible results. The basic configuration is adopted as a modification of the VGG16 structure [13].

The network consists of an input layer, three blocks, a flatten layer, and a single neuron at the output layer (Fig. 3.).

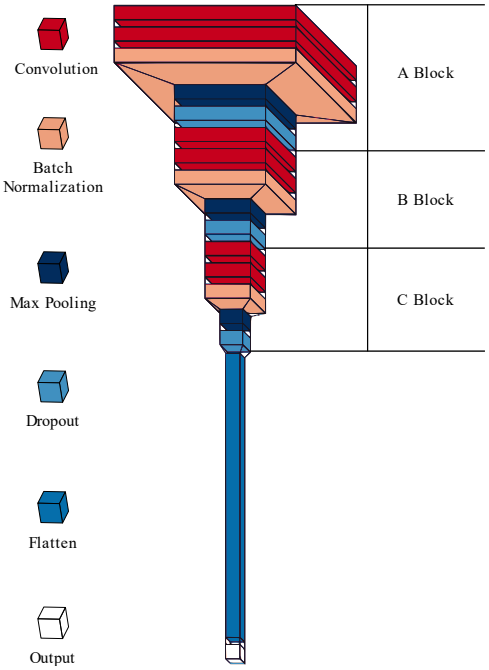

Fig. 3. Modified VGG16 neural network architecture.

The input of the network is a colored image obtained as a colored visualization of a data sequence, described in the preceding section. The input of the network has a shape of 64×64×3 representing the (height)×(width)×(color channels) of an image. This image size was selected so that both complex and simple gaze patterns could be clearly observed, without having a redundantly high resolution. Each of the block layers contains two convolutions, followed by a batch normalization layer, a single max pooling layer and a dropout layer (dropout rate of 0.5). The number of filters varies from block to block, with Block A having 16 filters, Block B having 32 filters and Block C having 64 filters. The strides of all convolution layers were (1,1) and the NAdam optimizer was used to train the network. All activation functions were ReLU except for the output neuron which had the sigmoid activation function. Several values for the kernel size of the convolution layers were evaluated separately (3, 5, 7), in combination with 2 different max pool kernel sizes (2, 4).

The entire implementation of the data processing and neural network design were performed in the Python 3 programming language. More specifically, the neural network implementation was performed using the tensorflow (version 2.8.0) library [14], and all the parameters for the neural network model that were not explicitly stated should be considered to be equal to the default values given in the functions of the mentioned libraries.

3) Neural network evaluation approaches

Each of the configurations of the neural network was evaluated on a data instance level as well as on the subject level. The data instance level signifies the classification results where each colored image is individually classified, and the average accuracy over all the data instances is displayed as a result. The subject level evaluation implies that all the class predictions for images belonging to a single subject, for a given color combination, are averaged, resulting in a total of 13 predictions per subject (one for each color configuration). Following this, the average prediction for the 13 color configurations is obtained at it is compared to the true class of the subject.

The evaluation was performed using leave-one-out subject-wise cross-validation (taking all the data from one subject as the test set and leaving the data from other subjects for the train set). The cross-validation contained 30 folds so that each subject's data could be in the test set. For each of the 30 folds, the train set was divided in a ratio of 80:20 in order to provide a true train set and a validation set used for early stopping.

III. Results and Discussion

The results of the cross-validation on an instance level are given in Table I and the results on a subject level are given in Table II.

It could be observed that the best overall achieved accuracy on the data instance level was 79% which was obtained for a 3-segment split per trial with the convolution kernel size of (7,7) and a max pooling kernel size of (2,2). The same configuration obtained the best results on a subject level, achieving an accuracy of 87%. The overall accuracy depends on the neural network configuration as well as the number of segments per trial but are consistently equal or greater than 70 % for both data instance evaluation and subject evaluation.

TABLE I. INSTANCE LEVEL RESULTS FOR ALL EXAMINED CONFIGURATIONS

Model parameters		Accuracy for n segments per trial		
Convolution kernel size	***Max pooling kernel size***	$n = 3$	$n = 5$	$n = 10$
(3,3)	(2,2)	0.72	0.74	0.73
(3,3)	(4,4)	0.73	0.75	0.73
(5,5)	(2,2)	0.75	0.73	0.72
(5,5)	(4,4)	0.76	0.74	0.75
(7,7)	(2,2)	0.79	0.76	0.76
(7,7)	(4,4)	0.78	0.76	0.74

TABLE II. SUBJECT LEVEL RESULTS FOR ALL EXAMINED CONFIGURATIONS

Model parameters		Accuracy for n segments per trial		
Convolution kernel size	***Max pooling kernel size***	$n = 3$	$n = 5$	$n = 10$
(3,3)	(2,2)	0.70	0.77	0.80
(3,3)	(4,4)	0.70	0.77	0.80
(5,5)	(2,2)	0.83	0.77	0.80
(5,5)	(4,4)	0.83	0.77	0.83
(7,7)	(2,2)	0.87	0.77	0.80
(7,7)	(4,4)	0.83	0.83	0.83

In addition to the displayed accuracies in Table I, a statistical analysis in the form of a Wilcoxon sum rank test was performed on the neural network predictions on all combined test sets, for each given configuration, between the dyslexic and control class. The achieved p value was lower than 0.001 for each configuration, indicating a clear separability between the classes given by the neural network predictions.

The best obtained results were achieved for the smallest number of segments per trial (3) with the highest observed convolution kernel size and the lower max pooling kernel size. This indicates that the neural network which is given more data at the input in a single image, with less instances (3 segments per trial) can obtain better results than the one with more instances, but less data per single image (5 and 10 segments per trial). The largest convolution kernel size indicates that the focus area of each layer should be larger, and that the data reduction factor (corresponding to the max pooling kernel size) should be low, allowing for the maximum amount of data to be propagated through the network. The model with this configuration could possibly be expected to exhibit signs of overfitting but there were no signs of overfitting in this case. Overfitting was evaded using the early stopping in each fold.

The achieved accuracy (87%) in this paper is less than the best achieved ones of 96.9% obtained by deep learning approach [7]. This is to be expected, as the text that was read within the experiment was written in Serbian, and the language used in the experiment in [7] was Swedish. The Serbian language has a much higher correspondence between phonemes and letters in comparison to Swedish, making dyslexia harder to diagnose. Some other research findings [5], [6], [15], [16] that automatically recognize dyslexia are based on feature extraction and traditional machine learning approaches that require more preprocessing in the form of eye tracking event detection, making their implementation more complex and sensitive to certain data anomalies.

The holistic approach with minimal data processing presented in this paper, in combination with a color-coded image representation of data has proven to be quite effective when paired with modern deep learning algorithms that stand out in image classification. The image format of the data also makes the data easier to visualize and interpret, while enabling the visual detection of potential data impurities, such as eye tracker errors, nonregistered gaze data or pauses during reading.

IV. CONCLUSION

In this paper, a method for the classification of dyslexia using 2D convolutional neural networks was presented. Several configurations of the trial segmentation were examined in combination with several values for the convolution kernel size and max pool kernel size.

The best achieved subject accuracy of 87% shows that the experiment design that includes eye tracking data during reading on a variety of color configurations can produce a higher accuracy than observing a single configuration. The best achieved accuracy on an instance level is 79% and the overall subject classification accuracy is higher by 8% because of the averaging of results for each subject over all the color configurations.

The directions for future work would include the analysis of the signals acquired by other sensor systems (heartrate, EDA, EEG) as well as an in-depth analysis of the eye tracking data. The combining of the dataset in this paper with another dataset that includes measurements of biometric signals during reading will also be considered.

REFERENCES

[1] C. Hulme and M. J. Snowling, "Learning to Read: What We Know and What We Need to Understand Better," *Child Dev. Perspect.*, vol. 7, no. 1, pp. 1–5, Mar. 2013, doi: https://doi.org/10.1111/cdep.12005.

[2] R. L. Peterson and B. F. Pennington, "Developmental dyslexia.," *Lancet (London, England)*, vol. 379, no. 9830, pp. 1997–2007, May 2012, doi: 10.1016/S0140-6736(12)60198-6.

[3] J. Miciak and J. M. Fletcher, "The Critical Role of Instructional Response for Identifying Dyslexia and Other Learning Disabilities," *J. Learn. Disabil.*, vol. 53, no. 5, pp. 343–353, Feb. 2020, doi: 10.1177/0022219420906801.

[4] G. T. Pavlidis, "Eye Movements in Dyslexia: Their Diagnostic Significance," *J. Learn. Disabil.*, vol. 18, no. 1, pp. 42–50, Jan. 1985, doi: 10.1177/002221948501800109.

[5] L. Rello and M. Ballesteros, "Detecting Readers with Dyslexia Using Machine Learning with Eye Tracking Measures," 2015. doi: 10.1145/2745555.2746644.

[6] M. Nilsson Benfatto, G. Öqvist Seimyr, J. Ygge, T. Pansell, A. Rydberg, and C. Jacobson, "Screening for Dyslexia Using Eye Tracking during Reading," *PLoS One*, vol. 11, no. 12, p. e0165508, Dec. 2016, [Online]. Available: https://doi.org/10.1371/journal.pone.0165508

[7] B. Nerušil, J. Polec, J. Škunda, and J. Kačur, "Eye tracking based

dyslexia detection using a holistic approach," *Sci. Rep.*, vol. 11, no. 1, p. 15687, 2021, doi: 10.1038/s41598-021-95275-1.

[8] F. Sultana, A. Sufian, and P. Dutta, "Evolution of Image Segmentation using Deep Convolutional Neural Network: A Survey," *Knowledge-Based Syst.*, vol. 201–202, p. 106062, 2020, doi: https://doi.org/10.1016/j.knosys.2020.106062.

[9] J. Naranjo-Torres, M. Mora, R. Hernández-García, R. J. Barrientos, C. Fredes, and A. Valenzuela, "A Review of Convolutional Neural Network Applied to Fruit Image Processing," *Appl. Sci.*, vol. 10, no. 10, 2020, doi: 10.3390/app10103443.

[10] S. Mohapatra, T. Swarnkar, and J. Das, "2 - Deep convolutional neural network in medical image processing," V. E. Balas, B. K. Mishra, and R. B. T.-H. of D. L. in B. E. Kumar, Eds. Academic Press, 2021, pp. 25–60. doi: https://doi.org/10.1016/B978-0-12-823014-5.00006-5.

[11] M. Browne and S. S. Ghidary, "Convolutional neural networks for image processing: an application in robot vision," in *Australasian Joint Conference on Artificial Intelligence*, 2003, pp. 641–652.

[12] T. Jakovljević, M. M. Janković, A. M. Savić, I. Soldatović, G. Čolić, T. J. Jakulin, G. Papa, and V. Ković, "The Relation between Physiological Parameters and Colour Modifications in Text Background and Overlay during Reading in Children with and without Dyslexia," *Brain Sci.*, vol. 11, no. 5, 2021, doi: 10.3390/brainsci11050539.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv Prepr. arXiv1409.1556*, 2014.

[14] Martín~Abadi *et al.*, "{TensorFlow}: Large-Scale Machine Learning on Heterogeneous Systems." 2015. [Online]. Available: https://www.tensorflow.org/

[15] 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, p. 20, 2019.

[16] T. Asvestopoulou, V. Manousaki, A. Psistakis, I. Smyrnakis, V. Andreadakis, I. M. Aslanides, and M. Papadopouli, "Dyslexml: Screening tool for dyslexia using machine learning," *arXiv Prepr. arXiv1903.06274*, 2019.