

RESEARCH ARTICLE

Eye-Tracking Image Encoding: Autoencoders for the Crossing of Language Boundaries in Developmental Dyslexia Detection

IVAN A. VAJS^{1,2}, GORAN S. KVAŠČEV¹, (Member, IEEE), TAMARA M. PAPIĆ³,
AND MILICA M. JANKOVIĆ¹, (Member, IEEE)

¹University of Belgrade - School of Electrical Engineering, 11120 Belgrade, Serbia

²Innovation Center, School of Electrical Engineering, University of Belgrade, 11120 Belgrade, Serbia

³Faculty of Technical Sciences, University Singidunum, 11010 Belgrade, Serbia

Corresponding author: Ivan A. Vajs (ivan.vajs@ic.etf.bg.ac.rs)

This work was supported by the Ministry of Education, Science and Technology Development of Serbia, Belgrade, Serbia, under Contract 451-03-68/2022-14/200103 and Contract 451-03-68/2022-14/200223.

ABSTRACT The negative influence of developmental dyslexia on academic performance is a well-documented and researched topic. Although the research focused on developmental dyslexia detection and evaluation is plentiful, the study designs vary to a great degree, making the exchange of obtained knowledge often difficult. This paper focuses on bridging the gap between different study designs by developing a machine learning based pipeline that was evaluated on two completely different eye-tracking datasets (training on one, testing on the other, and vice versa). One dataset included 30 subjects who read text written in Serbian on different color configurations and were tracked with a remote eye-tracker. The second dataset included 185 subjects who read text written in Swedish and recorded eye-tracking data using a goggle-based system. The data from both datasets were converted to grayscale images, using various time window configurations to parse the signals, and to plot the data in a 2D plane. The train images were used to train an Autoencoder neural network, and the images' reconstruction error was used to create features that describe each instance of both the training and test sets. The train feature set was used to train various machine learning algorithms, which were then evaluated on the testing feature dataset. A classification accuracy of 85.6% was obtained when testing on Serbian readers' data and 82.9% when testing on Swedish readers. The proposed pipeline was shown to be transferable between the datasets, despite many differences in the experiment design, showing potential in combining various eye-tracking dyslexia studies.

INDEX TERMS Autoencoder, classification, dyslexia, eye-tracking, machine learning, neural network.

I. INTRODUCTION

The influence that dyslexia can have on a child's academic performance was shown to be significant, often leading to low self-esteem [1], [2], [3]. Early treatment of dyslexia can greatly improve future academic performance and quality of life, making early diagnosis an important aspect in the field of dyslexia research [4], [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Li He¹.

Dyslexia diagnosis is conventionally performed using standardized tests, but considering that there are no firm agreements on the diagnosis procedure, introducing quantifiable and objective measures is of utmost importance [6]. A number of studies have used various experimental setups and biomedical features to objectify and quantify dyslexic tendencies. Some studies focus on measuring brain activities during reading, relying on functional magnetic resonance imaging [7], [8], [9], diffusion tensor imaging [8], [10] or electroencephalography (EEG) [11], [12], [13] to quantify

the differences between the dyslexic and non-dyslexic tendencies.

Aside from the cognitive load analysis, tracking of the visual sampling strategy, i.e. eye movements, is also often performed in dyslexia studies. Prado et al. reported one of the earliest eye-tracking dyslexia studies [14], but Rello et al. have reported being the first paper to perform dyslexia classification based on eye-tracking features [15]. They included 98 subjects in the study and used a support vector machine (SVM) to differentiate dyslexic and control subjects based on eye event features and participant characteristics. Multiple studies have taken a similar route, observing eye-tracking features and implementing machine learning (ML) algorithms to detect dyslexic tendencies [16], [17], [18], [19], [20], [21], [22]. The reading languages observed in these papers included Swedish, Spanish, Greek, Finnish, and French, all including eye-tracking measures and achieving various classification accuracies ranging from 80% to 96.6%. Some dyslexia studies use remote eye-tracking devices [20], [21], while others use wearable eye-trackers [16], [22]. Although quite different in experimental setups, these papers all prove that eye-tracking during reading can be successfully used for dyslexia detection. The research found in the literature also combines eye-tracking with other measurements, such as EEG, heart rate, and galvanic skin response (GSR) [23], [24] only EEG [25], and features acquired from behavioral measures [26].

Although features derived from biometric signals do show promising results, research involving a more holistic approach, utilizing biometric signals rather than features, is also present in dyslexia detection research. In [27], the authors analyze the dataset given in [16] and use one-dimensional convolutional neural networks (CNN) to detect dyslexics' eye movements. The authors observe the sequences of x and y gaze coordinates through time, implement several augmentation strategies and train a CNN to classify the signals. They obtain an accuracy of 96.6% and show that neural networks can indeed recognize dyslexic tendencies when presented with the entire eye-tracking signal.

Multiple different aspects of dyslexia research can be found in the literature, and the authors of this paper participated in several studies of dyslexia, following the trends in modern research. An initial study regarding the analysis of multimodal signals, including EEG, GSR, and eye-tracking measurements, was performed [24]. The study was done on native Serbian-speaking children and included an analysis of the influence of different background and overlay color configurations on dyslexic and non-dyslexic school-age children. The EEG, GSR, heart rate and eye-tracking measurements were used to extract features describing the cognitive load, emotional response, and visual sampling of the readers from both groups.

An in-depth analysis of the gathered eye-tracking data was performed in a later study and included an introduction of new eye-tracking features [28]. The novel features, focusing

on the spatial characteristics of fixations, were compared to the conventional eye-tracking features and have shown superior results in dyslexia classification. The evaluation was performed by using the features as inputs to various ML algorithms and using leave-one-out subject cross-validation. The best accuracy that could be obtained using conventional features was 85%, while a single newly introduced feature could provide an accuracy of 91%. The motivation behind these novel features was that the dyslexic group had more chaotic eye movements when fixating on a word in comparison to the control group. This could be observed in the x - y coordinate plane (the plane of the text display) as a scribble-like pattern, and the novel features quantified this using the number of self-intersections of gaze lines, as well as the estimation of the fractal dimension.

Adopting a more holistic approach, the spatial characteristics of the gaze have also inspired an image representation of eye-tracking data in another study [29]. By observing the same dataset, but using almost no signal processing, the gathered eye-tracking data was plotted in the x - y coordinate plane, color-coded, and converted to images. These images were then used to train and evaluate a deep two-dimensional CNN. The goal of the study was to use the same aspect of dyslexic tendencies as in [28] but to rely on the powerful pattern recognition capabilities of the CNN and an image representation of eye-tracking data.

Image processing represents a field with a wide range of applications and challenges, and exceptional results can be obtained using convolutional deep learning algorithms. From improving medical image quality, medical image segmentation, detecting surface defects, and providing super-resolution of photo-realistic images, these algorithms have been shown to have the immense potential [30]. One particular CNN architecture type that has stood out as particularly interesting to the authors of this paper was the CNN autoencoder (AE). This architecture consists of the encoder and decoder parts, which aim to transform the input to a compressed feature space and then reconstruct the original input from that compressed format. AEs have multiple applications, including classification [31], image fusion [32], image enhancement [33], noise reduction [34], and many more. One particular application of AEs that is of interest in this paper is anomaly detection [35]. AEs detect anomalies in data because they tend to reconstruct "typical" or more common data from the dataset better than the data that has a more complex or different pattern. This was used as a tool to create a pipeline introduced in this paper that could bridge the gap between different experimental scenarios for dyslexia eye-tracking studies. The focal point of the implemented pipeline is the fact that the spatial characteristics of gaze are more complex and specific for dyslexic readers, which could enable them to be distinguished from the reading patterns of controls by an AE.

Based on the spatial complexity principle in the method published in the author's previous study [28], this paper focuses on making the quantification of spatial

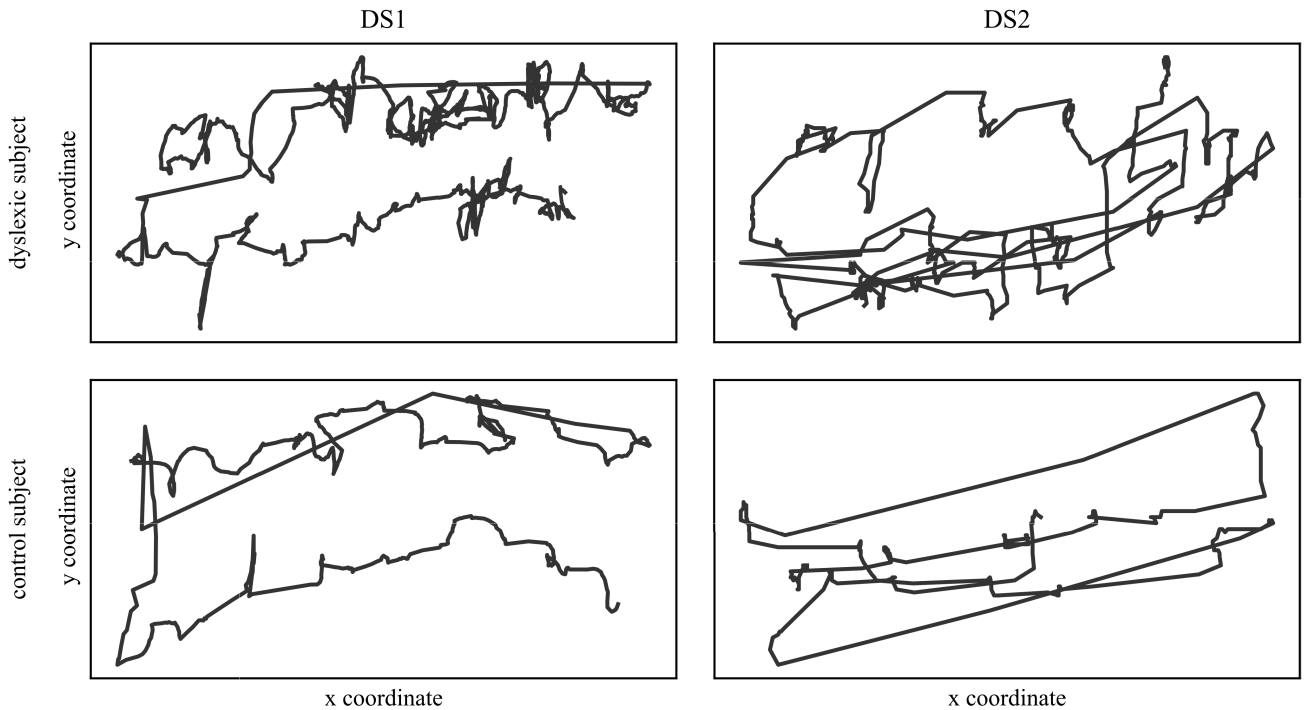


FIGURE 1. Examples of eye-movements displayed in the x-y (screen or paper) coordinate plane, for both DS1 (left column) and DS2 (right column). The top row displays the eye movements of dyslexic subjects, and the bottom row displays the eye movements of control subjects.

complexity more universal. The eye movements do not have to be segmented into fixations and saccades in order to apply the algorithm presented in this paper. The detection of fixations and saccades could be done in a multitude of ways which could be inconsistent across various experiments, which is why this paper proposes a method that relies only on the raw eye-tracking data. The pipeline for dyslexia detection suggested in this paper consists of standard processing and ML algorithms for the classification of dyslexic vs. control subjects, trained on data from one eye-tracking experiment, and successfully used on data from a completely different eye-tracking experiment (different regarding language and language orthography, type of used eye-tracker device, screen vs. paper text presentation, and amount of text displayed in the experiment). The suggested pipeline used an image representation of data and a CNN AE to quantify the spatial complexity of gaze and was shown to be universally applicable to two completely different sets of data.

II. METHOD

A. DATASETS

The research performed in this paper uses two different eye-tracking datasets gathered during reading. The first dataset (DS1) was acquired in [24], and the second dataset (DS2) was acquired in [16].

DS1 contains data gathered from 30 subjects (15 dyslexic and 15 control) with ages ranging from 7 to 13; 19 female and 11 male. The age distribution of subjects was the same

for the dyslexic and control group. The participants were seated in front of a computer monitor and keyboard, with their heads positioned on a chin rest. Each participant read (quiet reading) 13 text segments during the experiment. The segments (2-3 sentences) were extracted from the elementary school story called “Saint Sava and the villager without happiness.” Each segment was displayed on a different color configuration (colored background or overlay), with the first segment always being black text on a white background and the other segments having a pseudorandom order of background/overlay color configurations. In further text, the reading of each text segment for one subject will be called “a trial.” Even though 30 subjects were included in the initial study, each with 13 trials, the trials with insufficient focus on the displayed text (reading time less than 5 s) were excluded, resulting in a total of 378 trials included in this study. The subjects were monitored with a multimodal sensor hub [36], but for this study, only the eye-tracking data was used. An SMI RED-m 120 Hz portable remote eye-tracker (iMotions, Copenhagen, Denmark) was used to monitor eye movements (spatial resolution of 6 minutes of arc), and the data were acquired at 60 Hz.

DS2 contains data gathered from 185 (88 controls, 97 dyslexics) subjects, ages ranging from 9 to 10; 145 male and 40 female subjects. A goggle-based infrared corneal reflection system, Ober-2TM (Formerly Permobil Meditech, Inc., Woburn, MA), was used to track eye movements (spatial resolution of 5 minutes of arc), and the dataset that was available online had a sample rate of 50 Hz. During each trial, the subject was equipped with a pair of lightweight,

individually adjustable, head-mounted goggles. A chin and forehead rest was deployed to minimize head movements and stabilize the viewing distance at 45 cm. Calibration was performed manually prior to each recording by setting the signal gain of each axis separately for each eye. All subjects read the same text presented on a single page of white paper with high contrast, with the text consisting of 10 sentences.

The eye-tracking data of each trial from both datasets were clipped to remove the beginning and the end that contain no true reading eye movements. An example of the gaze displayed in the x-y (screen or paper) coordinate plane from both subject groups and both datasets is given in Fig. 1 (two rows of text were extracted for illustration purposes).

The clipping of the trials was done by visual inspection by two experts and excluded only the parts of the trials that clearly did not involve any reading activity. In total, this resulted in 378 clipped trials (193 controls, 185 dyslexics) from DS1 and 181 (88 controls, 93 dyslexics) clipped trials from DS2 (4 trials were excluded because neither the beginning nor end of the reading process could be observed in the data).

B. PROCESSING PIPELINE

The processing pipeline in this paper was designed to focus on the spatial complexity of gaze and to be as robust as possible. The initial step in the development of the pipeline was observing the data from the two available datasets and seeing if similar patterns emerged.

As shown in the examples given in Fig. 1, data from both datasets indicate that eye movements do have a more chaotic pattern in dyslexic subjects in comparison to the control ones. It was necessary to observe the trials not as a whole but as a series of smaller segments to ensure that the processing algorithm could pick up on the same chaotic spatial tendencies in both datasets.

The entire process of developing and training an algorithm for the detection of dyslexia on one dataset is given in Fig. 2. Step number ① is the extraction of the x and y coordinates over time from the trial data. This step implies the removal of blinks/missing data and the averaging of gaze coordinates from both eyes. Step number ② represents the division of each trial into many smaller segments and plotting them in the x-y plane. This step ensures that the complexity of the gaze is observed locally in the spatial domain. The data was parsed into segments using time windows (TW), and four different configurations were applied independently. The observed TW sizes were 0.5 s and 1 s, and the TW strides were 0.2 s and 0.5 s. All four combinations of these parameters were utilized to investigate whether a longer or shorter TW with more or less overlap would provide an adequate division of the trial into smaller segments. Taking into consideration that the reading duration varies, there will be a different number of images for each trial, which will be accounted for in the later processing steps.

Once the data is divided into time segments and plotted in the x-y coordinate plane, it is converted into grayscale

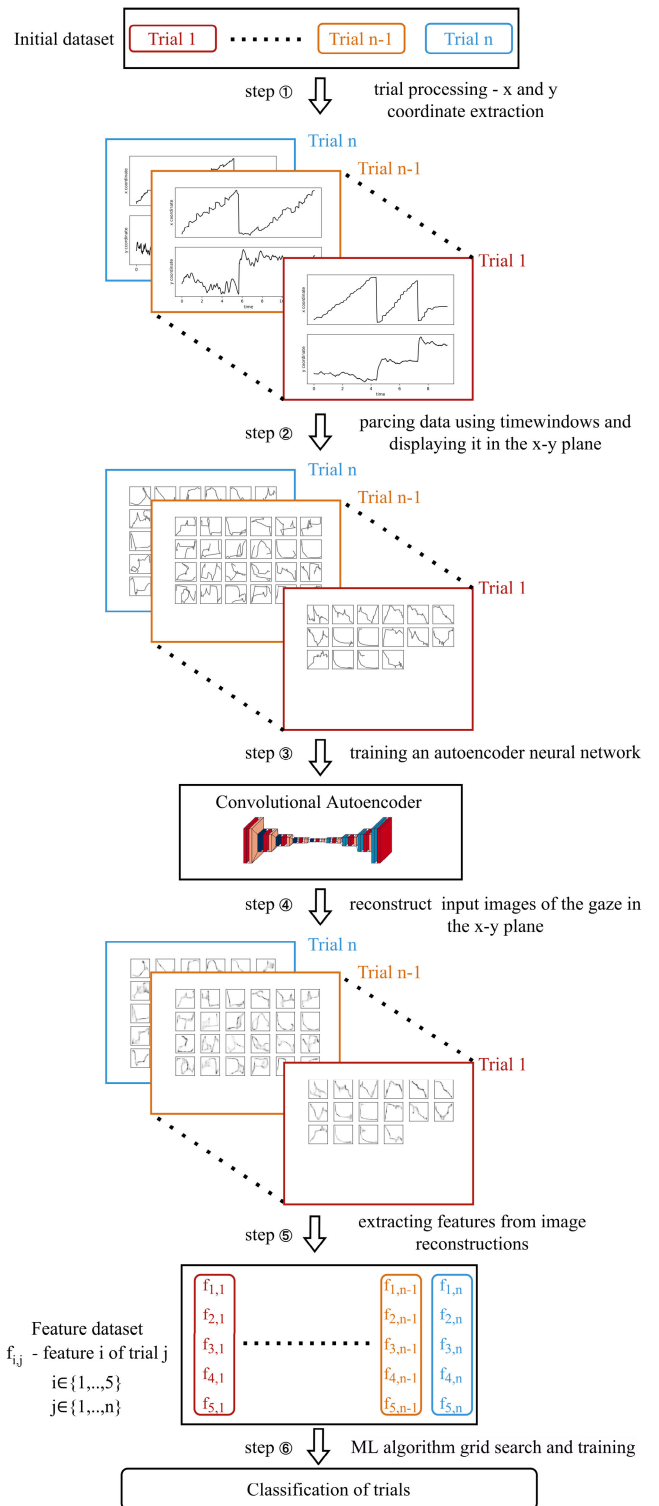


FIGURE 2. Processing pipeline of the algorithm for dyslexia detection.

images. The lines were plotted in black, with an opacity of 0.2 so that overlapping of lines could be observed. The images were resized into a 64×64 matrix, with the pixel values ranging from 0 to 255. The images were inverted, and the pixel values were normalized (divided by 255) so that the

background is black and the lines are a lighter shade of gray to facilitate easier numerical computation for the CNN AE.

The training of the CNN AE (step ③) was done on the resized and inverted images using a batch size of 64, the binary cross-entropy loss function, and the Adam optimizer. The binary cross-entropy loss was selected in spite of having a continuous range of pixel values to enable a way of label smoothing and to enable highlighting of the points where the lines intersect [37]. The same images were given to the input and the output of CNN AE, and the network was trained to reconstruct them. The training set was divided into 80% for true training and 20% for validation (for early stopping). The maximal number of epochs was set to 50.000 so that the training was always stopped by early stopping based on the loss on the validation set (with patience 20). The activation function for all layers was ReLu, except for the output layer which had the sigmoid activation function. The architecture of the CNN AE is given in Fig. 3.

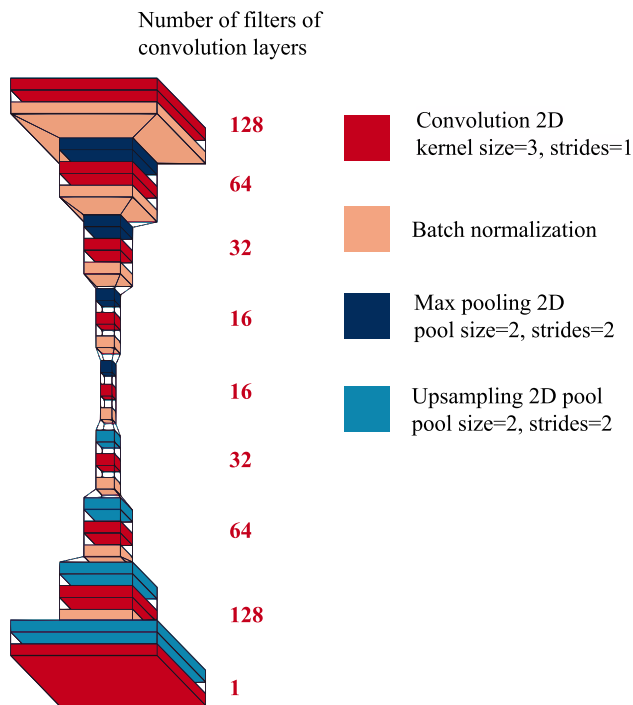


FIGURE 3. Convolutional Autoencoder (CNN AE) architecture.

Two scenarios for data selection were implemented for the training, one where only the control subjects' images were used for AE training and the other where all available data from one dataset was used to train the AE. The method that would be more conventional for AE training would be to use only the control images. However, considering the nature of the data, the fact that dyslexic readers do not struggle with reading at every single point in time but rather at certain intervals motivated the authors to try both scenarios in order to compare the results.

After the training of the AE, each trial can be represented by the reconstructions of the images created by plotting the x-y coordinates (step ④). These reconstructions were overall more similar to the original image for more simple patterns and less similar for more complex or specific patterns. An error metric calculated as the binary cross-entropy loss function between the original image and the reconstruction was used, describing each reconstruction with a single error value. Lower error values indicate a better reconstruction corresponding to a simple (more frequently occurred) gaze pattern, and the higher error values indicate a reconstruction that was not as successful, corresponding to more complex (less frequently occurred) gaze patterns.

With each image reconstruction described by the value of its error metric, the array of error metrics of a single trial is sorted, and 5 features were extracted by observing the differences between the 5 largest and 5 lowest values – step ⑤ of the pipeline, illustrated in Fig. 4. This results in each trial being represented by 5 features, with the features representing the ranges between the best and worst reconstructions of the trial. Dyslexic readers tend to have a less consistent gaze pattern and at times more complex gaze patterns, so the 5 features describing the range of gaze complexity are expected to have higher values for dyslexic subjects than the controls.

Each trial, represented by 5 features, is labeled according to its group (dyslexic or control), and the dataset of features and labels is used to train various ML algorithms. The algorithms considered were: logistic regression (LR), k-nearest neighbors (KNN), SVM, random forest (RF), and ada boost (AB). For each of the algorithms, a grid search for the best hyperparameter configuration was performed, using 5-fold cross-validation on the whole feature dataset. After the best hyperparameter was selected for each of the algorithms, they were trained on the entire feature dataset. This concludes step ⑥ and results in 5 trained classifiers.

The entire training process of the algorithms, consisting of the 6 mentioned steps, is developed on a single dataset and evaluated on the other one. When the evaluation is performed, the standard processing steps (①, ②, and ⑤) are performed the same way as they are during training. The AE is used to reconstruct the obtained input images of the dataset used for testing, but it is trained only on the training dataset. The same stands for the classification, trained classifiers are used on the testing dataset, but they are not trained on it. In this way, the entire pipeline is effectively developed on one dataset and evaluated on the other. Both DS1 and DS2 were used for training and testing, showing that the algorithm is useful and stable for both directions: training on DS1 and evaluation on DS2 and training on DS2 and evaluating on DS1.

The final metrics used for the evaluation of the trial classification were: Sensitivity (Se), Specificity (Sp), area under the receiver operating characteristic curve (AUROC), F1 score, and accuracy (ACC) [38].

Original images and reconstructions with error values in ‰, sorted according to the error values

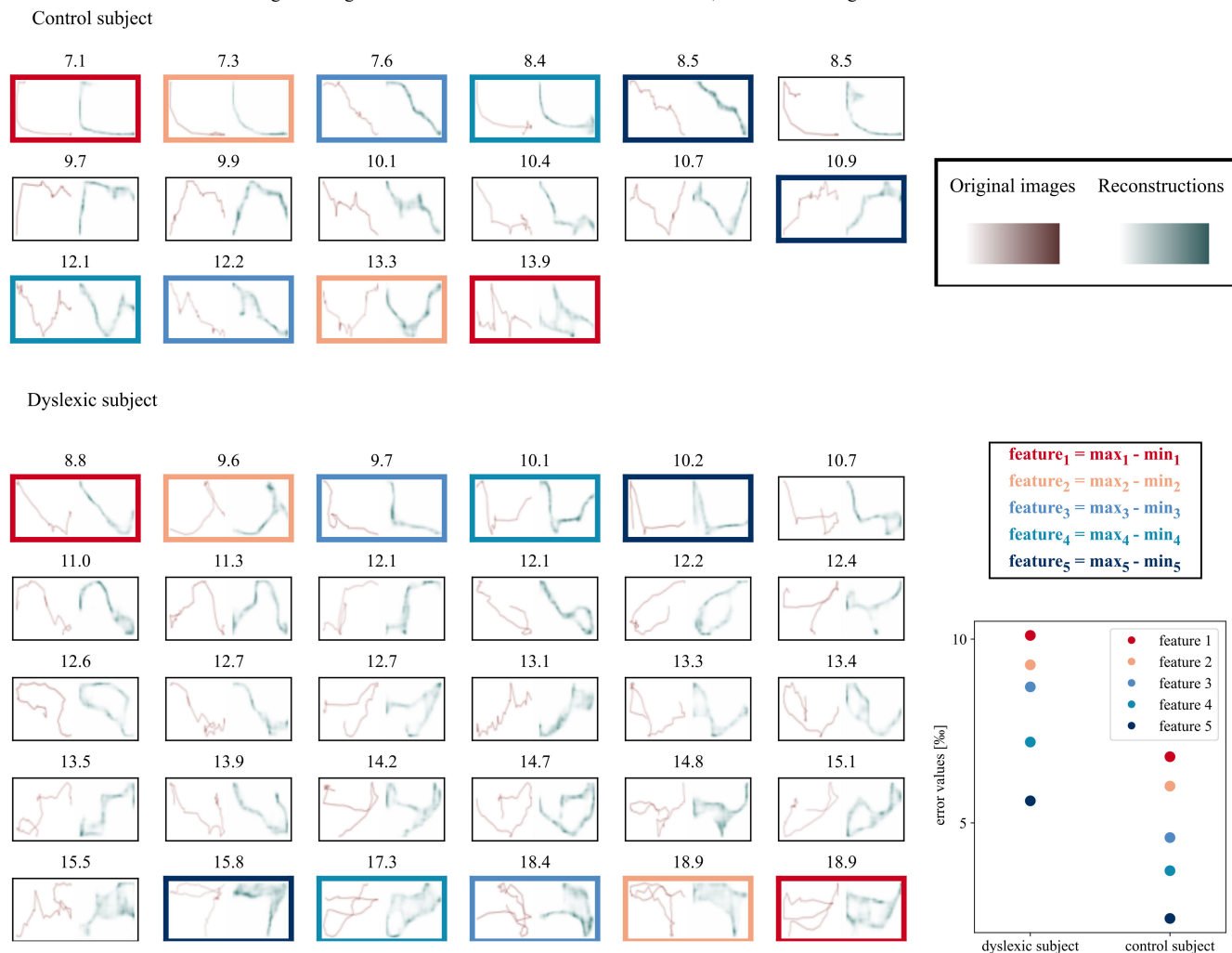


FIGURE 4. Feature extraction based on the image reconstruction errors, presented for a control and dyslexic subject.

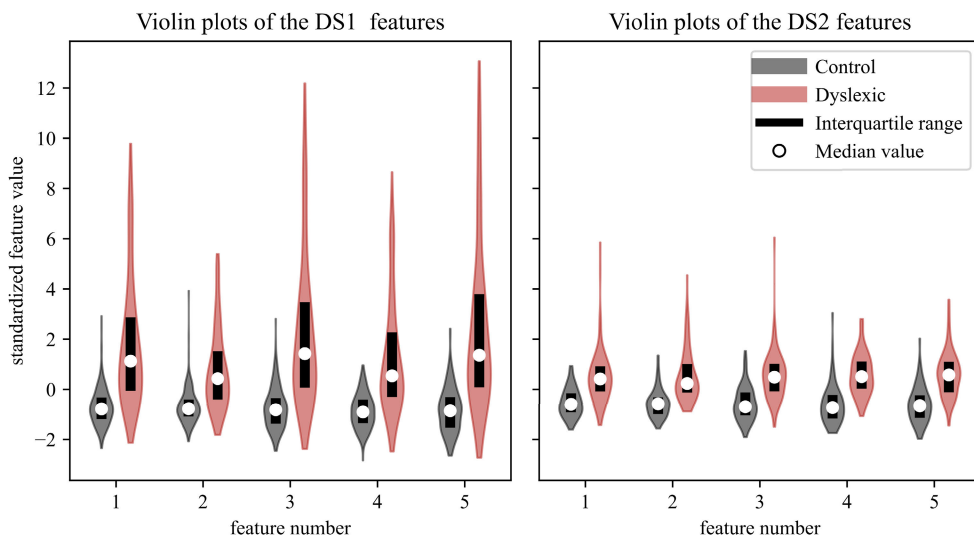


FIGURE 5. Extracted features for one evaluation scenario, displayed for the control and dyslexic group subject groups from DS1 and DS2.

TABLE 1. Evaluation results on DS2 (training on DS1).

TW size [s]	TW stride [s]	Dyslexics in AE training	ML algorithm	ACC [%]	F1 score [%]
0.5	0.5	No	LR	81.8	82.0
0.5	0.5	Yes	LR	80.7	80.2
0.5	0.2	No	LR	81.2	80.5
0.5	0.2	Yes	AB	81.8	81.8
1	0.5	No	LR	81.8	80.9
1	0.5	Yes	AB	81.2	80.0
1	0.2	No	LR	82.3	81.2
1	0.2	Yes	LR	82.9	81.7

All processing, ML algorithm implementation, and visualization of the data were done in the Python programming language, using the sklearn, tensorflow and matplotlib libraries [39], [40], [41], [42].

III. RESULTS AND DISCUSSION

The classification results are based heavily on the quality of the features that are given as inputs. To illustrate the features derived in the proposed algorithm, a single evaluation scenario was chosen (training on DS2, evaluation on DS1, TW size 0.5, TW stride 0.2, no dyslexic data in AE training), and the extracted features were displayed for both datasets for the two classes in Fig. 5. A clear separability between the classes can be seen, indicating that each feature individually does separate the classes effectively. A statistical analysis using the Mann-Whitney statistical test between the dyslexic and control groups was done, $p < 0.001$ (for each feature and evaluation scenario). The analysis was done on DS1 and DS2 separately. This, however, does not make the features redundant, as more features do provide a more stable prediction for each individual since the feature extraction process is sensitive to outliers.

Considering a large number of combinations of ML algorithms, TW configurations, AE training scenarios, and metrics, a simplified version of the results will be initially presented, giving only the ACC and F1-score of the best-performing ML algorithm for a given evaluation scenario. In Table 1, the results of the evaluation on the DS2 (training on DS1) are presented, and in Table 2 the results of the evaluation on DS1 (training on DS2) are presented.

The results achieved on the DS2 evaluation show an overall consistent ACC and F1 score achieved in all observed cases, and in most of them, LR was shown to be the superior algorithm. The best-obtained ACC and F1-score are 82.9% and 81.7%, respectively. As for the results achieved on the DS1 evaluation, they seem to be influenced by the evaluation scenario more than the DS2 evaluation. The best-achieved ACC is, however, slightly higher, equaling 85.6%. The F1 score for the best ACC is 84.9%, but the best overall F1 score is 85.2% and is obtained for a slightly lower ACC of 85.3%.

All the observed metrics are displayed for all ML algorithms for the best DS2 results and best DS1 results in Fig. 6 and Fig. 7, respectively.

TABLE 2. Evaluation results on DS1 (training on DS2).

TW size [s]	TW stride [s]	Dyslexics in AE training	ML algorithm	ACC [%]	F1 score [%]
0.5	0.5	No	LR	78.9	78.5
0.5	0.5	Yes	RF	84.1	83.4
0.5	0.2	No	LR	85.6	84.9
0.5	0.2	Yes	AB	85.3	85.2
1	0.5	No	LR	82.5	85.2
1	0.5	Yes	LR	82.5	82.0
1	0.2	No	LR	81.0	79.6
1	0.2	Yes	SVC	80.5	79.2

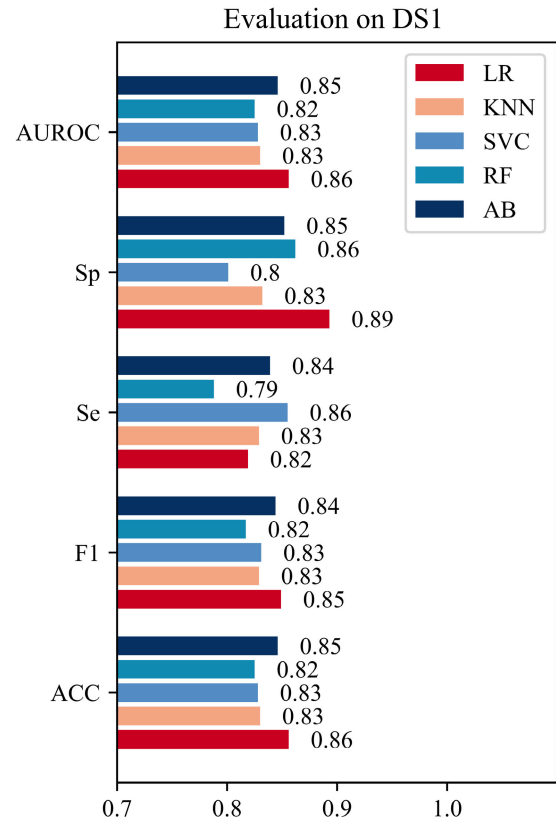


FIGURE 6. Overview of classification metrics for the best obtained results on DS2 in terms of TW and AE training scenarios, for all observed ML algorithms.

The results shown in Fig. 6 and Fig. 7. show different aspects of the DS2 and DS1 evaluations. Although the best-performing ML results were quite consistent for the DS2 evaluation in terms of different scenarios regarding the TW configuration and inclusivity of dyslexic subjects in AE training, there is quite a bit of variability in the performance between different ML algorithms. The best-performing algorithms are clearly LR and SVC, showing no difference in the ACC on the second decimal and having the best results in other metrics, with the SVC even having a higher specificity.

The other algorithms, although still providing better-than-random predictions, do not seem to perform quite as well. On the other hand, the results between various ML algorithms seem to be much more consistent on the DS1 evaluation. The variations between the algorithm results are much smaller than in the DS1 evaluation, despite the DS1 evaluation being

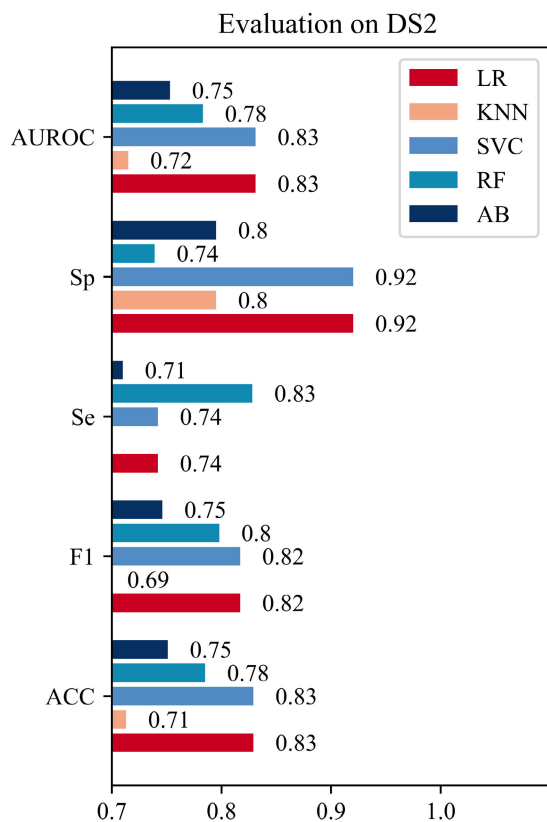


FIGURE 7. Overview of classification metrics for the best obtained results on DS1 in terms of TW and AE training scenarios, for all observed ML algorithms.

more sensitive to the TW configuration and dyslexic subject data inclusivity in AE training. This shows that various datasets might be influenced by different aspects of processing and ML algorithm selection, which could be dependent on a number of factors, such as the experiment protocol, number of participants, sensitivity of measuring instrument, etc.

The classification ACC found in the literature varies, ranging from 80% to 96.6% [15], [16], [17], [18], [19], [20], [21], [22], [28], [29], with the highest being 94% on DS1 [28] and 96.6% on DS2 [27], making the classification results obtained in this paper comparable to the ones in the literature, but not the highest. Having this in mind, all of the mentioned studies that focus on the detection of dyslexia using eye-tracking use only a single dataset, with all participants having the same native language, the same experimental protocol, instruments, and data structure. This study, however, aims to combine data from multiple studies that differ in many aspects. Focusing on the spatial characteristics of gaze, the goal of this study was not to present a state-of-the-art accuracy in dyslexia detection but to present an algorithm that could universally be applied to data from different studies. An algorithm that can be trained on one dataset and used on another has not yet been done in dyslexia eye-tracking research. In the presented research, one dataset [24] contains Serbian native readers, a remote eye-tracker, and a screen presentation of text on different color configurations, while

the other [16] contains Swedish native readers, a wearable eye tracker, and a presentation of text on a white piece of paper. Furthermore, the number of subjects (30 in DS1 vs. 181 in DS2) and the number of trials (378 in DS1 vs. 181 in DS2) in the two observed datasets are different as well. The amount of text displayed in the experiments is also different, with the Serbian dataset having two or three sentences while the Swedish dataset has 10 sentences. By observing the spatial characteristics, experiments with different amounts of text could be combined, as the temporal aspect of the data is not used in the proposed algorithm. Despite the differences, the processing pipeline trained on one dataset and evaluated on the other (and vice-versa) gives accuracies comparable to the ones in the literature. Different native languages of the readers can influence eye-tracking characteristics, as not all languages have the same depth of orthography. Still, the events that show the readers struggling with reading certain words were shown to be similar in nature for both Serbian and Swedish dyslexic readers. This indicates that dyslexic tendencies do have similar manifestations across different languages, at least in terms of the spatial characteristics of gaze. This shows that with an adequate combination of ML and traditional processing, it is possible to cross the language barrier and enable dyslexia detection across multiple languages and study designs.

IV. CONCLUSION

A novel method for dyslexia detection was presented in the paper. The method focuses on quantifying the spatial complexity of gaze segments by plotting them in the plane of the text and representing them as grayscale images. The images were used to train an AE, and the complexity of the gaze displayed in each image is quantified as the reconstruction error of the image. The ranges of reconstruction errors were extracted from each reading trial and used as input features for various ML algorithms. The method was evaluated using two datasets, training the models on one dataset and evaluating them on the other. The best-achieved accuracies were 85.6% when evaluated on DS1 and 82.9% when evaluated on DS2.

The developed method does not depend on the parsing of gaze data (into fixations and saccades) and does not rely on any temporal aspects. It was also evaluated on two datasets with completely different experimental protocols, subject pools, and gaze-tracking instruments. The obtained results, comparable to ones found in the literature, show that the method is robust and can successfully be used to cross many boundaries between different dyslexia studies. Future work would include an analysis with more datasets, implementation of various types of AE, and the investigation of the methods' sensitivity to a lower spatial resolution of the used eye-tracking data.

ACKNOWLEDGMENT

The authors would like to thank M.Sc. EE Milica Badža Atanasijević and Ph.D. Vladislava Krsmanović for their useful advice and suggestions.

REFERENCES

- [1] E. Singer, "Coping with academic failure, a study of Dutch children with dyslexia," *Dyslexia*, vol. 14, no. 4, pp. 314–333, Nov. 2008, doi: [10.1002/dys.352](https://doi.org/10.1002/dys.352).
- [2] E. Lithari, "Fractured academic identities: Dyslexia, secondary education, self-esteem and school experiences," *Int. J. Inclusive Educ.*, vol. 23, no. 3, pp. 280–296, Mar. 2019, doi: [10.1080/13603116.2018.1433242](https://doi.org/10.1080/13603116.2018.1433242).
- [3] F. Polychroni, K. Koukoura, and I. Anagnostou, "Academic self-concept, reading attitudes and approaches to learning of children with dyslexia: Do they differ from their peers?" *Eur. J. Special Needs Educ.*, vol. 21, no. 4, pp. 415–430, Nov. 2006, doi: [10.1080/08856250600956311](https://doi.org/10.1080/08856250600956311).
- [4] G. Schulte-Körne, "The prevention, diagnosis, and treatment of dyslexia," *Deutsches Ärzteblatt Int.*, vol. 107, no. 41, pp. 26–718, Oct. 2010, doi: [10.3238/arztebl.2010.0718](https://doi.org/10.3238/arztebl.2010.0718).
- [5] A. van der Leij, "Dyslexia and early intervention: What did we learn from the Dutch dyslexia programme?" *Dyslexia*, vol. 19, no. 4, pp. 241–255, Nov. 2013, doi: [10.1002/dys.1466](https://doi.org/10.1002/dys.1466).
- [6] 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, doi: [10.1109/ACCESS.2021.3062709](https://doi.org/10.1109/ACCESS.2021.3062709).
- [7] M. A. Lobier, C. Peyrin, C. Pichat, J.-F. Le Bas, and S. Valdois, "Visual processing of multiple elements in the dyslexic brain: Evidence for a superior parietal dysfunction," *Frontiers Human Neurosci.*, vol. 8, p. 479, Jul. 2014.
- [8] Y. G. Chimeno, B. G. Zapirain, I. S. Prieto, and B. Fernandez-Ruanova, "Automatic classification of dyslexic children by applying machine learning to fMRI images," *Bio-Medical Mater. Eng.*, vol. 24, no. 6, pp. 2995–3002, 2014, doi: [10.3233/BME-141120](https://doi.org/10.3233/BME-141120).
- [9] O. L. Usman, R. C. Muniyandi, K. Omar, and M. Mohamad, "Gaussian smoothing and modified histogram normalization methods to improve neural-biomarker interpretations for dyslexia classification mechanism," *PLoS ONE*, vol. 16, no. 2, Feb. 2021, Art. no. e0245579, doi: [10.1371/journal.pone.0245579](https://doi.org/10.1371/journal.pone.0245579).
- [10] M. Vandermosten, L. Cuyenen, J. Vanderauwera, J. Wouters, and P. Ghesquière, "White matter pathways mediate parental effects on children's reading precursors," *Brain Lang.*, vol. 173, pp. 10–19, Oct. 2017, doi: [10.1016/j.bandl.2017.05.002](https://doi.org/10.1016/j.bandl.2017.05.002).
- [11] M. Arns, S. Peters, R. Breteler, and L. Verhoeven, "Different brain activation patterns in dyslexic children: Evidence from EEG power and coherence patterns for the double-deficit theory of dyslexia," *J. Integrative Neurosci.*, vol. 6, no. 1, pp. 175–190, Mar. 2007, doi: [10.1142/S0219635207001404](https://doi.org/10.1142/S0219635207001404).
- [12] G. F. González, M. J. W. Van der Molen, G. Žarić, M. Bonte, J. Tijms, L. Blomert, C. J. Stam, and M. W. Van der Molen, "Graph analysis of EEG resting state functional networks in dyslexic readers," *Clin. Neurophysiol.*, vol. 127, no. 9, pp. 3165–3175, Sep. 2016, doi: [10.1016/j.clinph.2016.06.023](https://doi.org/10.1016/j.clinph.2016.06.023).
- [13] C. Spironelli, B. Penolazzi, and A. Angrilli, "Dysfunctional hemispheric asymmetry of theta and beta EEG activity during linguistic tasks in developmental dyslexia," *Biol. Psychol.*, vol. 77, no. 2, pp. 123–131, Feb. 2008, doi: [10.1016/j.biopsycho.2007.09.009](https://doi.org/10.1016/j.biopsycho.2007.09.009).
- [14] C. Prado, M. Dubois, and S. Valdois, "The eye movements of dyslexic children during reading and visual search: Impact of the visual attention span," *Vis. Res.*, vol. 47, no. 19, pp. 2521–2530, Sep. 2007, doi: [10.1016/j.visres.2007.06.001](https://doi.org/10.1016/j.visres.2007.06.001).
- [15] L. Rello and M. Ballesteros, "Detecting readers with dyslexia using machine learning with eye tracking measures," in *Proc. 12th Int. Web Conf.*, 2015, pp. 1–8, doi: [10.1145/2745555.2746644](https://doi.org/10.1145/2745555.2746644).
- [16] 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, Dec. 2016, Art. no. e0165508, doi: [10.1371/journal.pone.0165508](https://doi.org/10.1371/journal.pone.0165508).
- [17] 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, doi: [10.1016/j.cmpb.2020.105538](https://doi.org/10.1016/j.cmpb.2020.105538).
- [18] A. J. Prabha, R. Bhargavi, and B. Harish, "An efficient machine learning model for prediction of dyslexia from eye fixation events," *New Approaches Eng. Res.*, vol. 10, no. 6, pp. 171–179, Aug. 2021, doi: [10.9734/bpi/naer/v10i11914D](https://doi.org/10.9734/bpi/naer/v10i11914D).
- [19] J. P. Appadurai and R. Bhargavi, "Eye movement feature set and predictive model for dyslexia: Feature set and predictive model for dyslexia," *Int. J. Cogn. Informatics Nat. Intell.*, vol. 15, no. 4, pp. 1–22, 2021.
- [20] 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, doi: [10.1016/j.array.2021.100087](https://doi.org/10.1016/j.array.2021.100087).
- [21] T. Asvestopoulou, V. Manousaki, A. Psistakis, I. Smyrnakis, V. Andreadakis, I. M. Aslanides, and M. Papadopoulou, "DysLexML: Screening tool for dyslexia using machine learning," 2019, *arXiv:1903.06274*.
- [22] A. E. El Hmimdi, L. M. Ward, T. Palpanas, and Z. Kapoula, "Predicting dyslexia and reading speed in adolescents from eye movements in reading and non-reading tasks: A machine learning approach," *Brain Sci.*, vol. 11, no. 10, p. 1337, Oct. 2021, doi: [10.3390/brainsci11101337](https://doi.org/10.3390/brainsci11101337).
- [23] T. Jakovljević, M. M. Janković, A. M. Savić, I. Soldatović, P. Todorović, T. J. Jakulin, G. Papa, and V. Ković, "The sensor hub for detecting the developmental characteristics in reading in children on a white vs. colored background/colored overlays," *Sensors*, vol. 21, no. 2, p. 406, 2021, doi: [10.3390/s21020406](https://doi.org/10.3390/s21020406).
- [24] T. Jakovljević, M. Jankovic, A. M. Savić, I. Soldatović, and G. Colić, "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, p. 539, 2021, doi: [10.3390/brainsci11050539](https://doi.org/10.3390/brainsci11050539).
- [25] B. Weiss, Á. Nárai, and Z. Vidnyánszky, "Lateralization of early orthographic processing during natural reading is impaired in developmental dyslexia," *NeuroImage*, vol. 258, Sep. 2022, Art. no. 119383, doi: [10.1016/j.neuroimage.2022.119383](https://doi.org/10.1016/j.neuroimage.2022.119383).
- [26] B. Weiss, J. Szalma, and Z. Vidnyánszky, "Data-driven detection of developmental dyslexia: A machine learning approach based on behavioral and eye-movement features," *PsyArXiv*, Jun. 2022, doi: [10.31234/osf.io/qasnc](https://doi.org/10.31234/osf.io/qasnc).
- [27] 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, Aug. 2021, doi: [10.1038/s41598-021-95275-1](https://doi.org/10.1038/s41598-021-95275-1).
- [28] I. Vajs, V. Ković, T. Papić, A. M. Savić, and M. M. Janković, "Spatiotemporal eye-tracking feature set for improved recognition of dyslexic reading patterns in children," *Sensors*, vol. 22, no. 13, p. 4900, 2022, doi: [10.3390/s22134900](https://doi.org/10.3390/s22134900).
- [29] I. Vajs, V. Ković, T. Papić, A. M. Savić, and M. M. Janković, "Dyslexia detection in children using eye tracking data based on VGG16 network," in *Proc. 30th Eur. Signal Process. Conf. (EUSIPCO)*, 2022, pp. 1601–1605.
- [30] L. Jiao and J. Zhao, "A survey on the new generation of deep learning in image processing," *IEEE Access*, vol. 7, pp. 172231–172263, 2019, doi: [10.1109/ACCESS.2019.2956508](https://doi.org/10.1109/ACCESS.2019.2956508).
- [31] H. Patel and K. P. Upla, "A shallow network for hyperspectral image classification using an autoencoder with convolutional neural network," *Multi-media Tools Appl.*, vol. 81, no. 1, pp. 695–714, 2022, doi: [10.1007/s11042-021-11422-w](https://doi.org/10.1007/s11042-021-11422-w).
- [32] H. Zhang, H. Xu, and X. Tian, "Image fusion meets deep learning: A survey and perspective," *Inf. Fusion*, vol. 76, pp. 323–336, Dec. 2021, doi: [10.1016/j.inffus.2021.06.008](https://doi.org/10.1016/j.inffus.2021.06.008).
- [33] S. Khamekhem Jemmi, M. A. Souibgui, Y. Kessentini, and A. Fornés, "Enhance to read better: A multi-task adversarial network for handwritten document image enhancement," *Pattern Recognit.*, vol. 123, Mar. 2022, Art. no. 108370, doi: [10.1016/j.patcog.2021.108370](https://doi.org/10.1016/j.patcog.2021.108370).
- [34] A. Pawar, "Noise reduction in images using autoencoders," in *Proc. 3rd Int. Conf. Intell. Sustain. Syst. (ICISS)*, 2020, pp. 987–990, doi: [10.1109/ICISS49785.2020.9315908](https://doi.org/10.1109/ICISS49785.2020.9315908).
- [35] N. Shvetsova, B. Bakker, I. Fedulova, H. Schulz, and D. V. Dylov, "Anomaly detection in medical imaging with deep perceptual autoencoders," *IEEE Access*, vol. 9, pp. 118571–118583, 2021, doi: [10.1109/ACCESS.2021.3107163](https://doi.org/10.1109/ACCESS.2021.3107163).
- [36] T. Jakovljević, M. M. Janković, A. M. Savić, I. Soldatović, P. Todorović, T. J. Jakulin, G. Papa, and V. Ković, "The sensor hub for detecting the developmental characteristics in reading in children on a white vs. colored background/colored overlays," *Sensors*, vol. 21, no. 2, p. 406, 2021, doi: [10.3390/s21020406](https://doi.org/10.3390/s21020406).
- [37] E. Gordon-Rodriguez, G. Loaiza-Ganem, G. Pleiss, and J. P. Cunningham, "Uses and abuses of the cross-entropy loss: Case studies in modern deep learning," in *Proc. NeurIPS Workshops*, vol. 137, 2020, pp. 1–10. [Online]. Available: <https://proceedings.mlr.press/v137/gordon-rodriguez20a.html>
- [38] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: An overview," 2020, *arXiv:2008.05756*.
- [39] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA, USA: CreateSpace, 2009.

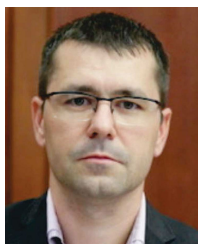
- [40] F. Pedregosa, G. Varoquau, A. Gramfort, V. Michel, and B. Thirion, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Jan. 2011.
- [41] M. Abadi. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: <https://www.tensorflow.org/>
- [42] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, May/June. 2007.



IVAN A. VAJS received the B.Sc. and M.Sc. degrees in electrical engineering from the School of Electrical Engineering, University of Belgrade, Belgrade, Serbia, in 2019 and 2020, respectively, where he is currently pursuing the Ph.D. degree.

He is also a Junior Research Assistant with the Innovation Center, School of Electrical Engineering, University of Belgrade. His research interests include signal processing in the field of biomedical engineering, data science, and machine learning.

He has participated in two projects that were part of the Innovation Vouchers Program of the Innovation Fund of the Republic of Serbia and the Project Development of Science and Technology of the Ministry of Education, Science and Technological Development of the Republic of Serbia. He has also participated in the Project AVANTES of the Program for Development of Projects in the field of artificial intelligence of the science fund of the Republic of Serbia and a Horizon 2020 Project—L4MS. He is also working on his Ph.D. in the field of biomedical signals and machine learning regarding dyslexia. He is also working on machine learning methods in low-cost air quality sensor calibration.

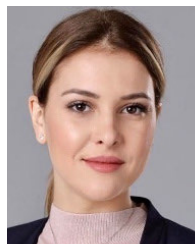


GORAN S. KVAŠČEV (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the School of Electrical Engineering (ETF), University of Belgrade, Belgrade, Serbia, in 2000, 2005, and 2012, respectively.

He joined the School of Electrical Engineering, University of Belgrade, where he is currently an Associate Professor, teaching courses in control systems theory, industrial process control, neural networks, and deep learning. He is also the Head

of Signals and Systems Department. Since 2000, he has been realized several projects in field of industrial process control, fault detection, HVAC, signal processing, and informatics in medicine. He has participated in several TEMPUS, FP7, EUREKA, and WUS projects.

Prof. Kvaščev is a member of the Serbian Chamber of Engineers and a licensed engineer on field of automation and control and IT systems.



TAMARA M. PAPIĆ received the Ph.D. degree from Jožef Stefan International Postgraduate School, Ljubljana. Her doctoral dissertation was multidisciplinary research in the field of sensor technologies, dedicated to the development of the sensor hub for the detection of the color influence on reading in children with reading difficulties, such as dyslexia.

She was a Scholar of the AD Futura Fund of the Republic of Slovenia. She is currently an Assistant Professor with the Faculty of Technical Sciences, University Singidunum, and the Co-Founder of health tech startup Baby FM which won three international prizes (Local Imagine IF Program, Innovation Forum-Cambridge, Raising starts-Swiss Government and Science-Technology Park Belgrade, and Star Tech Program). During her multidisciplinary work, she surrounded herself with prominent professors from Slovenia and Serbia in order to help the educational process in children through the early detection and prevention of Dyslexia. Her research work has been presented at several international conferences and published in three international journals, like *Sensors*, *PLOS One*, and *Brain Sciences*.



MILICA M. JANKOVIĆ (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the School of Electrical Engineering (ETF), University of Belgrade, Belgrade, Serbia, in 2003, 2008, and 2014, respectively.

She is currently an Associate Professor and the Head of the Laboratory for Biomedical Engineering and Technology, ETF. As a principal investigator on behalf of ETF, she participated in two projects of the Innovation Fund Republic of Serbia (one Collaborative Grant Scheme for Research and Development Organizations and Private Sector Enterprises and one Innovation Voucher), Erasmus+ K2 Strategic Partnership (ITASDI), and one PokreniNauku Project. On behalf of ETF, she was the Local Coordinator of the CEEPUS Network of Image Processing, Information Engineering and Interdisciplinary Knowledge Exchange. She participated in three projects of Ministry of Education, Science and Technological Development, two bilateral international projects, one H2020 project (DIH-HERO) and one project of the Science fund of the Republic of Serbia (ForNextCobot). She has published in 22 publications in indexed scientific journals. Her research interests include medical instrumentation, measurement, data acquisition, and medical imaging. Her papers were honored with awards, including the Etran 2005 and 2014, the Telfor 2013, the Balcan Congress of Nuclear Medicine 2015, the Second Prize on International Medical Olympiad 2017, and the Best Paper at Experts in Stone Disease Conference 2022.

...