MASARYK UNIVERSITY

FACULTY OF INFORMATICS

# Designing Visual Representations of Eye-Tracking Data for Improving Ensemble-Based Dyslexia Classification

Bachelor's Thesis

ADAM DOSTÁL

Brno, Spring 2025

MASARYK UNIVERSITY

FACULTY OF INFORMATICS

# Designing Visual Representations of Eye-Tracking Data for Improving Ensemble-Based Dyslexia Classification

Bachelor's Thesis

ADAM DOSTÁL

Advisor: doc. RNDr. Jan Sedmidubský, Ph.D.

Department of Machine Learning and Data Processing

Brno, Spring 2025


FACULTAS ARTIS INFORMATICÆ
UNIVERSITAS MASARYKIANA

# Declaration

Hereby I declare that this paper is my original authorial work, which I have worked out on my own. All sources, references, and literature used or excerpted during elaboration of this work are properly cited and listed in complete reference to the due source.

During the preparation of this thesis, I utilized the following artificial intelligence tools:

- **Grammarly** to improve my writing style and correct grammar and spelling mistakes.

I declare that I used these tools in accordance with the principles of academic integrity. I checked the content and took full responsibility for it.

Adam Dostál

**Advisor:** doc. RNDr. Jan Sedmidubský, Ph.D.

# Acknowledgements

I want to express my sincere gratitude to my thesis supervisor, doc. RNDr. Jan Sedmidubský, Ph.D., for his expert guidance, valuable advice, and continuous support throughout the development of this thesis. This thesis was also supported by the Technology Agency of the Czech Republic (TL05000177 – Diagnostics of dyslexia using eye-tracking and AI).

# Abstract

Dyslexia is a learning disorder that affects around 5 % to 17.5 % of people worldwide and is primarily characterized by word reading difficulties, which can lead to poor academic performance. Thus, early detection of dyslexia is necessary. Current approaches for dyslexia detection aim at classifying eye-tracking data, which are recorded primarily during text-reading tasks. Those data are usually transformed into compact representations such as high-dimensional vector features or image visualizations, which are then used to train machine-learning classifiers. This thesis aims to improve the classification accuracy of current models by proposing new visualizations of raw eye-tracking data and derived characteristics, such as fixations and saccades. To further enhance classification accuracy, the proposed visualizations are divided into sub-images. For each sub-image, an independent deep neural network is trained, and the outputs are finally ensembled by majority voting. Experimental results show that the proposed visualizations, along with the division of visualizations into sub-images and the ensemble of deep neural network outputs, achieve a high accuracy of up to 92.86 %.

# Keywords

dyslexia, classification, ResNet-18, majority voting, visualizations, eye tracking, eye movement

# Contents

# List of Tables

# List of Figures

# 1 Introduction

Dyslexia is a learning disorder characterized by severe word decoding and spelling difficulties. These symptoms result in slower reading fluency and challenges with text comprehension. Beyond poor academic outcomes, dyslexia is also associated with the social (e.g., behavioral problems) and emotional development of dyslexic individuals [1, 2]. Early and more precise detection of dyslexia is crucial for the further development of reading skills of dyslexics. This not only improves their reading skills but also enhances their academic performance [1, 3, 4].

Dyslexia detection is a complex process because symptoms can vary widely among individuals. This process typically utilizes a multidisciplinary approach and involves various tasks such as reading assessments, leveraging the expertise of psychologists, teachers, and social workers [5, 6]. Dyslexia impacts around 5 % to 17.5 % of people worldwide [2]. One of the promising research directions is to detect dyslexia automatically by capturing *eye-tracking data* during text-reading tasks and classifying such data using machine learning principles [3].

Eye-tracking data provide valuable insights into reading and how an individual interacts with written text. Eye-tracking data can be captured using an eye-tracking device, with a capturing frequency ranging from 30 to 2000 Hz [7]. *Raw* eye-tracking data are represented as a sequence of 2D points, where the $x$ and $y$ coordinates indicate the position on the monitor screen where the individual's eye is looking at [3].

Such raw data can further be processed to extract meaningful high-level *gaze events*, where the most common ones are *fixations* and *saccades*. A fixation is a period of time during which the eye remains relatively still, focusing on a specific point, and represents a cluster of raw data points that are close to each other. A saccade is a quick eye movement that shifts gaze from one fixation to a consecutive fixation. A saccade in the direction of reading is called a forward saccade (progression), and in the opposite direction is called a backward saccade (regression) [3, 8, 9]. An illustration of raw data, fixations, and saccades is shown in Figure 2.1.

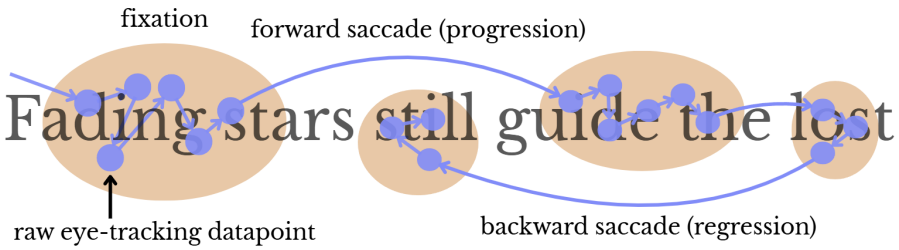



**Figure 1.1:** Illustration of raw eye-tracking data shown as blue circles, fixations represented as orange circles, and both types of saccades depicted as blue lines connecting raw datapoints.

Other characteristics can be derived from fixations and saccades, such as fixation duration, saccade length, or regression frequency. Characteristics can be represented in several ways (e.g., numerical statistics as vectors), and these representations can be classified using various machine learning techniques, such as *k*NN, SVM, or neural networks. One of the most promising approaches is the use of neural networks, due to their ability to capture complex patterns [10]. Neural networks can process input characteristics represented as numerical statistics in the form of vectors or as graphical representations of captured eye-tracking data in the image format [3, 9].

This thesis aims to improve the classification accuracy of state-of-the-art methods [3] by introducing new types of visualizations of eye-tracking data. Furthermore, the proposed visualizations are divided into sub-images, and an independent deep neural network is trained for each sub-image. The results are then ensembled together using majority voting, which leads to the improvement of classification accuracy.

This thesis begins with a review of the current state of the art in Chapter 2. Chapter 3 describes the dataset used for the proposed methods, while Chapter 4 introduces and explains the proposed visualization methods. Chapter 5 then presents the results of applying these methods to ResNet-18. Finally, Chapter 6 summarizes the evaluated methods' findings and outlines potential directions for future research.

# 2 Related Work

The detection of dyslexia through eye-tracking data collected during reading tasks has been the subject of numerous studies [11]. This thesis provides a brief overview of existing classification methods and datasets, emphasizing the most promising ones relevant to this thesis. This chapter summarizes commonly used data representations and machine learning techniques.

## 2.1 Most-Related Approaches

Among the first attempts for dyslexia detection using machine learning techniques was a study by Rello et al. [12]. This study involved 97 participants (48 dyslexic) aged from 11 to 54. They assessed performance on 12 different text-reading tasks and achieved an accuracy of up to 80.18 % using Support Vector Machines (SVM).

In a study by Nilsson Benfatto et al. [13], they achieved up to 95.6 % accuracy using SVMs with Recursive Feature Elimination (RFE). Their study is based on data collected from 185 children (97 high-risk) aged from 9 to 10 on one text-reading task. The dataset from this study was also used by Nerušil et al. [14], where they achieved an accuracy of 96.6 % using a Convolutional Neural Network (CNN).

Another feature-based representation approach was conducted by Vajs et al. [15]. This research collected data from 30 Serbian children (15 dyslexic) on reading tasks. For dyslexia recognition, they have used several machine learning techniques such as $k$-Nearest Neighbours ($k$NN), random forests, and SVMs, achieving an accuracy of up to 94 %.

Another research was done by Černek in his Master's thesis [16], where he achieved an accuracy of up to 85.88 % on the feature-based representation approach using Multi-Layer Perceptrons (MLP) and up to 92 % on a visual-based representation approach using Residual Neural Networks (ResNet). Both approaches used an ensemble method based on majority voting across four reading tasks. The research involved 35 children between the ages of 9 and 10 (13 were dyslexic).

A recent contribution to the area of dyslexia detection is a study by Sedmidubsky et al. [3], where they achieved an accuracy of up to 90 % using the feature-based representation approach with MLP and up to 88 % with the visual-based representation using ResNets. The research involved three reading tasks and included 70 Czech children aged 9 to 10 years (35 were dyslexic).

## 2.2 Reading Tasks

The most common task for testing participants is the reading task, which typically consists of meaningful text [11]. Usually, in the reading task, each word in every sentence is treated as a rectangle, also known as an *area-of-interest* (AOI). An example is shown in Figure 2.1, where red rectangles represent the AOIs. AOIs do not have to correspond exactly to words. In general, they may represent any area of any shape that participants should focus on. Further, characteristics for one task can be extracted globally from the whole task, considering all AOIs, or individually, concentrating on specific AOIs.


**Figure 2.1:** Example of AOIs on a custom English sentence.

## 2.3 Data Representations

All mentioned approaches [3, 12, 13, 14, 15, 16] use feature-based data representation rather than raw eye-tracking data. This is because the amount of raw eye-tracking data is significantly larger than the number of extracted characteristics. Some approaches [3, 16] also use visual-based representations, which may visualize gaze events, such as fixations or saccades, in combination with some of the extracted characteristics, such as fixation duration.

### 2.3.1 Feature-based Representations

The raw eye-tracking data are low-level. They contain $x$ and $y$ coordinates and timestamps from when the record was captured. Thus, looking into more high-level statistics from gaze events (e.g., fixations or saccades) can provide the required level of detail. Those statistics can be derived globally from one task, including all AOIs, or individually per AOI.

Global statistics typically include a count of gaze events, primarily for fixations, progression, and regression saccades. Other global features mainly reflect high-level statistics about changes in eye movements, typically by summarizing the length and amplitude of saccades using metrics such as mean, median, 75th, and 25th quartiles. Similar statistics can also be made for fixations by summarizing their durations and $x$ and $y$ dispersions. Those statistics can also be extracted individually per AOI [16].

In summary, feature-based representations are defined as high-dimensional vectors of real numbers selected from defined statistics. In the study by Nilsson Benfatto et al. [13], they managed to extract 48 features using RFE. In Černek's thesis [16], he extracted up to 46 features, while in the study by Sedmidubsky et al. [3], they managed to extract up to 450 features for one task.

### 2.3.2 Visual-based Representations

Feature-based representations include aggregated high-level characteristics such as means, medians, or quartiles. However, they overlook individual characteristics like fixation position or fixation duration, along with their changes over time. Instead of representing features as vectors of real numbers, some of them, particularly fixations, can be represented graphically in the image format. This is accomplished by visualizing $x$ and $y$ coordinates of fixation as a circle with a fixed size and a single color.

One of the simplest visualizations based on fixations is a heatmap. In heatmap representation, the intensity of the color indicates the level of attention given to specific areas of a stimulus (visual content presented to the participants). Warmer colors signify areas that received more attention (more fixations), while cooler colors indicate areas

with less attention (fewer fixations) [9]. Figure 2.2 shows an example of fixation visualization using a heatmap.

Recent approaches [4, 16] came with a visualization method based on fixations for reading tasks. The visualization image has a fixed size, where each fixation is represented as an ellipsoid, where the center of the ellipsoid is determined by the fixation's $x$ and $y$ coordinates. The fixation dispersion in the $x$ and $y$ axes determines the ellipsoid's width and height. The normalized value of fixation duration determines the color of the ellipsoid. Figure 2.3 shows this fixation visualization method.

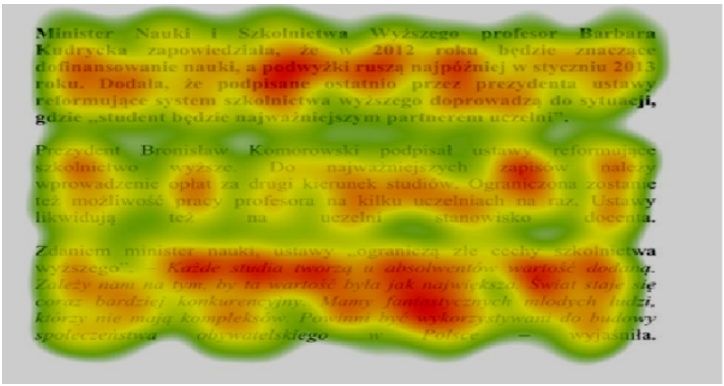


**Figure 2.2:** Example of representing fixations as a heatmap [17].

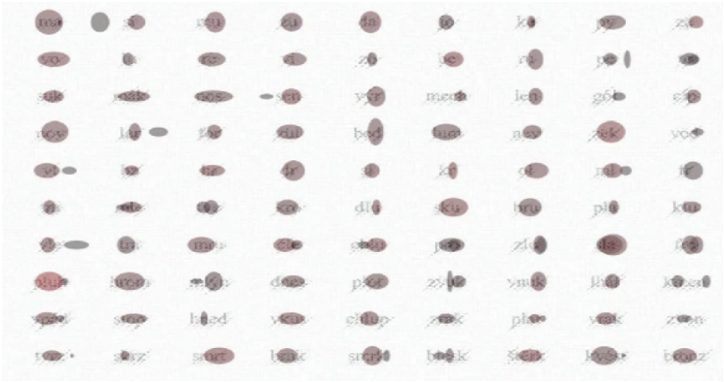


**Figure 2.3:** Example of fixation visualization, where each fixation is represented as an ellipsoid [3].

## 2.4 Classification Approaches

Each type of data representation is suited to specific machine-learning approaches for classification. A *k*-Nearest Neighbors (*k*NN) similarity search, Support Vector Machines (SVM), and Multi-Layer Perceptron (MLP) architectures are suitable for classifying feature-based representations [11]. In contrast, using Convolutional Neural Networks (CNN) or Residual Neural Networks (ResNet) for visual-based representations is more appropriate, since they offer better classification accuracy on images than MLPs [18, 19].

### 2.4.1 *k*NN Classification

One of the most commonly used methods for machine learning is the supervised *k*-Nearest Neighbors algorithm. In this algorithm, each feature vector represents points in the $n$-dimensional space $\mathbb{R}^n$. The feature vector can be derived from a feature-based representation or extracted using a trained neural network, for example, as the content of the last hidden network layer. The concept of *k*NN is based on the idea that points belonging to the same class tend to be close to each other. To classify a new point, the algorithm calculates the distances to all training points and selects the $k$ nearest ones. The class that appears most frequently among these neighbors is then assigned to the new point. The most common method for measuring the distance between points is Euclidean distance [20, 21].

### 2.4.2 MLP Classification

The backbone of neural networks is the multi-layer perceptron, where neurons are organized into layers, with each neuron connecting to every neuron in the previous layer. Each neuron applies an activation function to introduce non-linearity into the model. The first layer is the input layer, which contains $n$ neurons corresponding to the dimension of the feature vector. The output layer often uses the softmax function across all neurons in one layer to convert output scores into probabilities that sum to one. The layers between the input and output are known as hidden layers, which play a crucial role in learning patterns in the data [20].

One of the MLP architectures developed for dyslexia classification was introduced by Sedmidubsky et al. [3]. This architecture consists of an input layer, three hidden layers, and an output layer. The input layer receives a 450-dimensional feature vector. Each subsequent hidden layer reduces the dimensionality by half (the first hidden layer has 225 neurons, the second 112, and the third 56). The softmax activation function is utilized for the output layer to perform two-class classification.

### 2.4.3 ResNet Classification

While MLPs are better suited for feature-based representations, Convolutional Neural Networks are more effective for image classification based on visual representations. CNNs are more advanced than MLPs and typically achieve higher classification accuracy [18]. However, for deeper neural networks, it might be better to use Residual Neural Networks (ResNet-18, ResNet-50) as they are more efficient, robust, and generally achieve better accuracy than CNNs [10, 19, 22].

A pre-trained network that accepts a fixed-size image as input can be utilized for small datasets, such as the one used in this thesis. Those pre-trained networks are usually trained on general domain images and then fine-tuned on the generated visual representations. This also requires modifying the original classification layer (e.g., 1,000 classes in a model trained on a dataset of general domain images) to accommodate only two classes for dyslexia detection [3].

## 2.5 Summary of Advantages/Disadvantages

Each classification approach comes with its own set of advantages and disadvantages. Feature-based classification methods, like *k*NNs and MLPs, are less computationally intensive than deep neural networks. However, they are not suitable for visual-based classification tasks. In contrast, deep neural networks like ResNets are essential for these tasks despite their significantly higher computing power requirements.

This thesis contributes to improving the classification accuracy of state-of-the-art methods using ResNets by designing new visualizations containing raw eye-tracking data, fixations, and saccades.

Building upon the framework established in this study [3], where visualizations are put into ResNets as an entire image of a fixed size. However, due to the large size of these visualizations, compressing them into a smaller fixed size can lead to data loss. To prevent data loss, smaller images are necessary. For this reason, this thesis divides the entire image into several smaller sub-images, where each sub-image is used to fine-tune an individual ResNet model. The results are then ensembled using majority voting.

The image-splitting technique generates sub-images whose size is closer to the required input size for ResNet. This technique keeps more details after compression. The results from ResNet for individual sub-images may provide a detailed lookup on which sub-image is important and which is not, according to their accuracy. The ones with less accuracy may contain some noise or may be neutral (no difference between dyslexic and non-dyslexic participants). This may be considered during ensembling, where only some sub-images with higher accuracy may be ensembled.

# 3 ETDD70 Dataset

The ETDD70 dataset [3] used in this thesis was developed by a research team from the Faculty of Arts of Masaryk University and the Faculty of Informatics of Masaryk University[1]. The dataset includes raw eye-tracking data and various event-based characteristics based on dyslexic and non-dyslexic participants while completing reading tasks defined in Section 3.2 and visual-differentiation tasks defined in Section 3.3. Data from visual-differentiation tasks were not published in this dataset but were recorded simultaneously on the same participants. The dataset is publicly available at: `https://zenodo.org/records/13332134`

## 3.1 Participants

The dataset was collected from 70 participants, consisting of 35 dyslexic individuals (19 females and 16 males) and 35 non-dyslexic individuals (19 females and 16 males). All participants were elementary school pupils aged 9 to 10 years, corresponding to the 4th grade of elementary school [3].

## 3.2 Text-Reading Tasks

The dataset focuses on three reading tasks, all in the Czech language.

- *Task 1* is a reading task containing a grid of 90 syllables (9 rows and 10 columns). The complexity of syllables increases with each row, where the first row has easier syllables, while the last one has more complex syllables. Participants had to read syllables from left to right, row by row, starting from the top left [3].
- *Task 2* is a reading task that contains a meaningful story about a young boy watching a squirrel from his window. The text has in total seven lines and six consecutive sentences [3].

1. The dataset was part of the project by the Technology Agency of the Czech Republic (TL05000177 – Diagnostics of dyslexia using eye-tracking and AI).

- *Task 3* is a reading task that contains a fictional text with meaningless words. The text consists of 7 text lines and 15 artificial sentences [3].

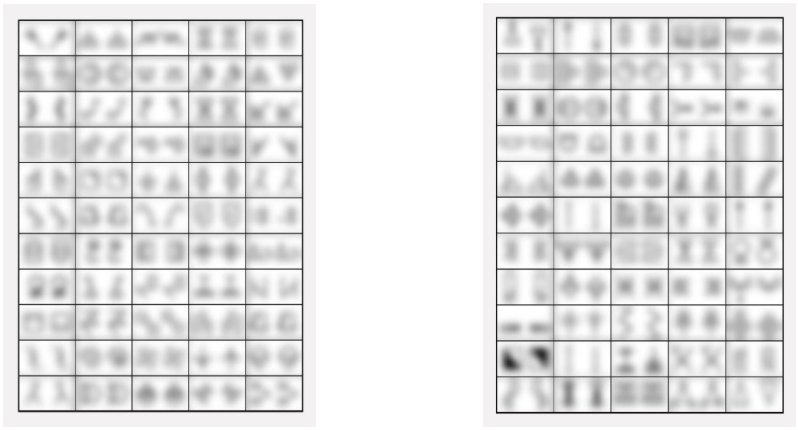
## 3.3 Visual-Differentiation Tasks

A visual differentiation task presents pairs of symbols that may match or not. The difference between non-matching symbols can be identified based on their top-bottom or right-left positioning. Children need to distinguish those detailed differences. Otherwise, if they struggle with visual differentiation, it may lead to confusion with letters and numbers differing in detail (such as m–n, k–h, 6–9, 4–7) or may result in confusion of letters and numbers differing in top-bottom positioning (t–j, b–p, 6–9) or right-left positioning (d–b, 6–9) [23].

*Task 4.1* and *Task 4.2* represent matching and non-matching pairs of symbols in a $11 \times 5$ grid (11 rows and 5 columns). Non-matching symbols include symbols that differ in detail (such as missing circles or shorter lines), and symbols that differ in top-bottom and right-left positioning. The objective for participants in Task 4.1 and Task 4.2 was to go through all pairs of symbols, starting from the top left pair and continuing row by row to the bottom right pair, and click on all pairs that differ. Visualization of Task 4.1 and Task 4.2 is shown in Figure 3.2. Since the visualizations are blurred at the request of the authors, an illustration of the symbol is shown in Figure 3.1. The only difference between Task 4.1 and Task 4.2 is the use of different symbols.

(**a**) (**b**)

**Figure 3.1:** Illustration of similarity differences on figure (a) showing a house with and without a chimney and right-left positioning differences on (b) showing half-ellipsoids with a small line.



(**a**) Task 4.1 (**b**) Task 4.2

**Figure 3.2:** Blurred visualization of Task 4.1 and Task 4.2.

## 3.4 Provided Data

The raw eye-tracking data were collected using the SMI-RED 250 remote eye-tracking device, which has a sampling rate of 250 Hz. Those data were obtained during reading-based and visual-based tasks presented on a monitor with a refresh rate of 60 Hz and a resolution of $1680 \times 1050$ pixels for each child. All children were tested at the same place at Masaryk University to preserve the same conditions for measurement [3].

The raw eye-tracking data were processed to extract gaze events, primarily fixations and saccades. The fixations were identified using an I2MC fixation detection algorithm [24]. This algorithm effectively manages data with varying noise levels, including cases of data loss. Consequently, this algorithm is well-suited for data recorded from children. The saccades were defined from fixations, where each saccade is a transition between consecutive fixations [3].

# 4 Proposed Visualization Representations

This thesis focuses on generating new visualizations based on visualizing raw eye-tracking data, fixations, and saccades. ResNet-18 with pre-trained weights was utilized to classify these visualizations since it provides better efficiency and robustness than other deep neural networks for image classification [10, 19, 22]. However, using pre-trained ResNet-18 presents several problems outlined in Section 4.1, primarily compression of visualization of fixed-size input for ResNet-18, resulting in data loss. The subsequent sections (from Section 4.2 to Section 4.5) describe generated visualizations. Section 4.6 defines the different versions of majority voting that were used. Last Section 4.7 describes the framework used for implementation and the hyperparameters that can be tuned.

## 4.1 Problem Definition

The input for this thesis consists of raw eye-tracking data represented as a sequence of 2D points. From raw eye-tracking data, fixations are derived as a sequence of 2D points, fixation durations, and dispersions in both $x$ and $y$ axes. Associated with fixations are saccades, defined as a sequence of pairs of 2D points, where one 2D point is the start of the saccade and the second 2D point is the end of the saccade. These sequences of points are then used for visualizations in the form of an image of specific dimensions (Task 1 has dimensions of 927 $(width) \times$ 682 $(height)$ pixels, Task 2 and Task 3 have $978 \times 532$ pixels, Task 4.1 and Task 4.2 have $552 \times 723$ pixels).

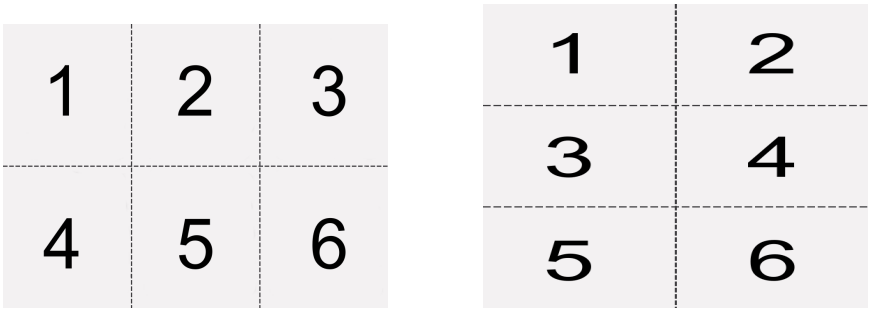
This thesis uses a ResNet-18 model with pre-trained weights trained on general domain images sized at $224 \times 224$ pixels. According to this, visualizations are compressed to the fixed size of $224 \times 224$ pixels and serve as input to a ResNet-18 model, which is fine-tuned on these visualizations to perform classification. The model outputs a probability value $p \in [0, 1]$, representing the estimated likelihood that a participant exhibits dyslexia-related characteristics.

The issue with the current approach is compressing an entire visualization image into smaller dimensions as input to the ResNet-18 model, leading to data loss. Using visualizations as an entire image

comes with another issue. By focusing on specific AOIs, there may be areas with AOIs where participants show no significant differences between dyslexic and non-dyslexic groups. This can lead to lower classification accuracy, as the model might be affected by irrelevant areas. Noise from these areas with uninformative AOIs may overshadow patterns that genuinely reflect the differences between dyslexic and non-dyslexic groups.

The objective of this thesis is to address these issues by dividing the full-size image into six distinct sub-images, ensuring that each sub-image has dimensions close to the desired size of $224 \times 224$ pixels. For reading tasks, the full-size images are vertically divided into three sub-images using one third of the *width* and then horizontally into two sub-images using one half of the *height*, resulting in a total of six sub-images. For visual-differentiation tasks, the division process is reversed, horizontally divided into three sub-images using one third of the *height* and vertically into two sub-images using one half of the *width*, resulting again in six sub-images. Each sub-image is used to fine-tune a separate, independent instance of the pre-trained ResNet-18 model, which evaluates the participant's performance in the specific sub-image area. The results are then ensembled together using majority voting. The numbering of sub-images is systematically assigned to correspond with the specific regions shown in Figure 4.1 for both reading and visual-differentiation tasks.

The division method was used on already existing visualization approaches provided by Sedmidubsky et al. [3] and also on proposed generated visualizations that are described in the following sections (raw eye-tracking data visualizations in Section 4.2 and Section 4.3, fixations with saccades in Section 4.5).

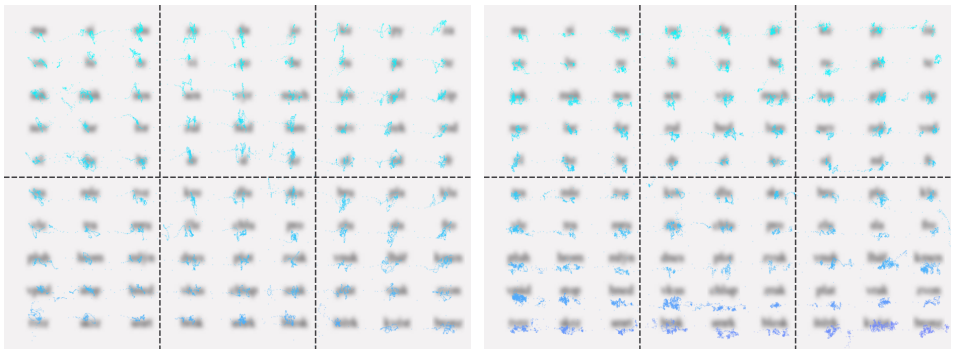



(**a**) Reading tasks (**b**) Visual-differentiation tasks

**Figure 4.1:** Sub-image numbering for reading and visual-differentiation tasks.

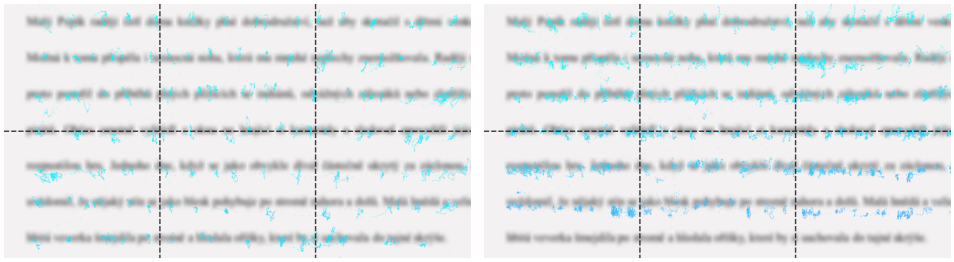
## 4.2 Method RD-Idx – Raw Data Index Color

This method visualizes raw eye-tracking data by plotting circles based on $x$ and $y$ coordinates recorded over time. Each raw datapoint is represented as a circle of fixed size (i.e., diameter) of 1.86 pixels. The datapoints index determines the circle's color (i.e., timestamp order), transitioning from cool blue for the minimal index values through purple to pink for the highest index values. The color mapping is normalized based on the dataset's [3] highest number of raw datapoints for each task individually. Figure 4.2 shows visualizations of this method on all tested tasks with a comparison of dyslexic and non-dyslexic participants. The dotted lines in the Figure 4.2 indicate the sub-image division as shown in the Figure 4.1.



(**a**) Non-dyslexic (**b**) Dyslexic

**Task 1**

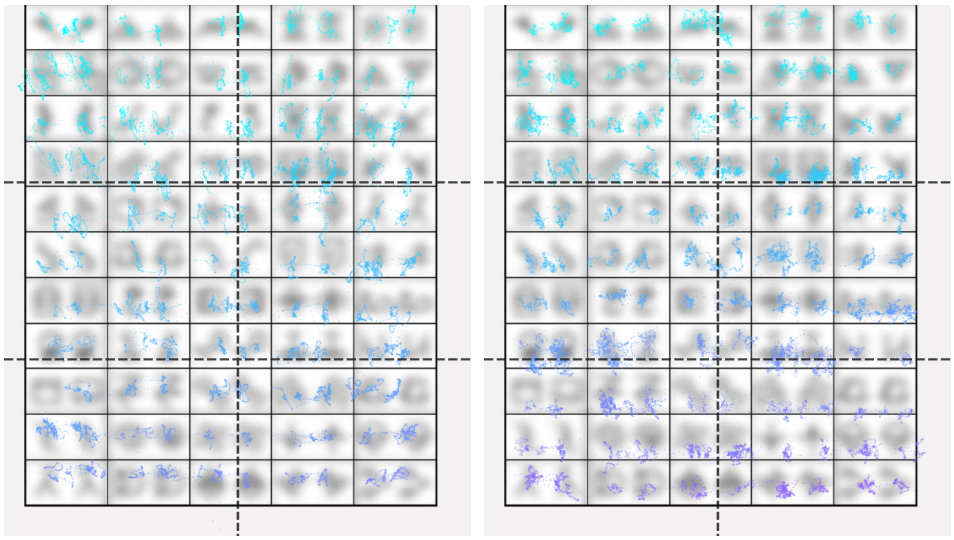


(**a**) Non-dyslexic (**b**) Dyslexic

**Task 2**

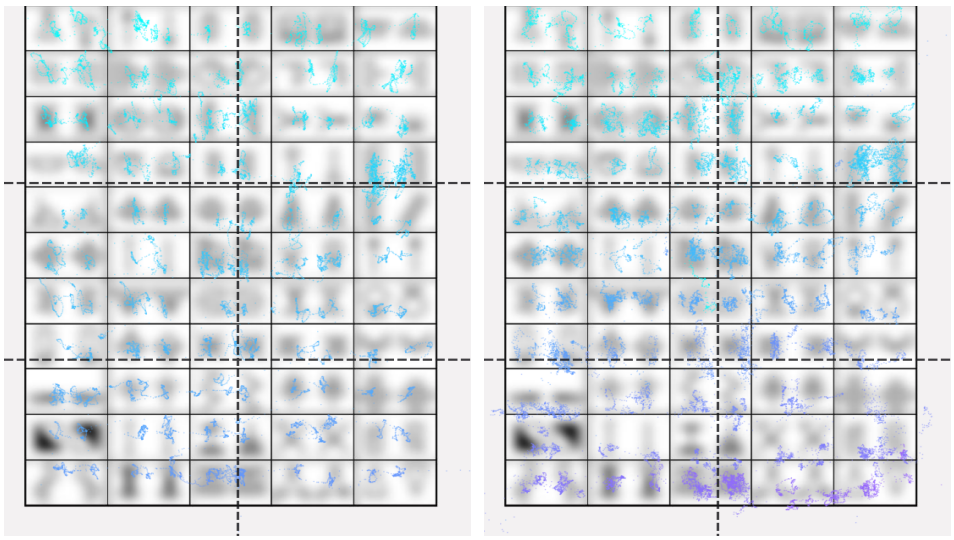
(**a**) Non-dyslexic (**b**) Dyslexic

**Task 3**



(**a**) Non-dyslexic (**b**) Dyslexic

**Task 4.1**



(**a**) Non-dyslexic (**b**) Dyslexic

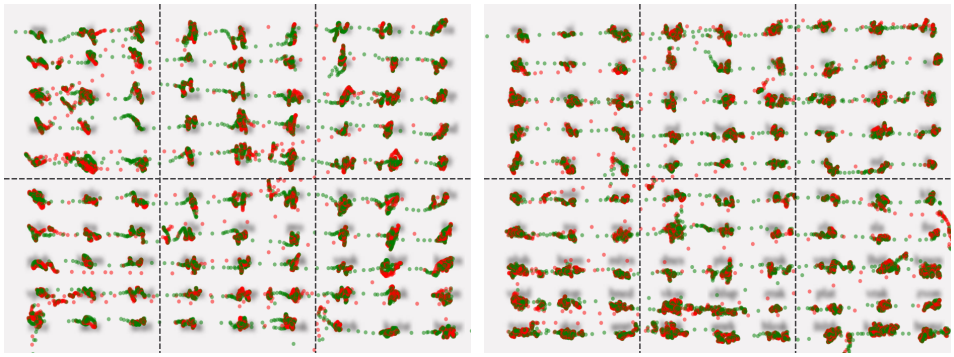
**Task 4.2**

**Figure 4.2:** Comparison of Method RD-Idx on non-dyslexic and dyslexic participants on all tasks.

## 4.3 Method RD-Sac – Raw Data Saccade Color

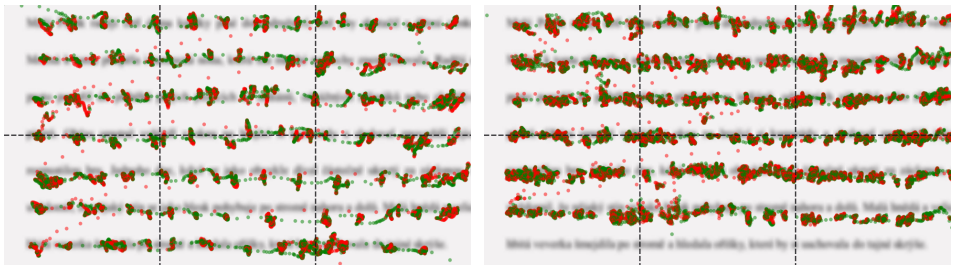
This method visualizes raw eye-tracking data by plotting circles with color based on saccade characteristics, instead of timestamp order as in the previous method described in Section 4.2. The circles have a fixed size (i.e., diameter) of 9.28 pixels. The green color indicates forward saccades, while the red color represents regression saccades (i.e., saccades in the opposite reading direction). This method is visualized in Figure 4.3 on all tested tasks, comparing dyslexic and non-dyslexic participants.

Using saccades as colors might be beneficial because participants with dyslexia tend to have more regression saccades than non-dyslexic participants [25].



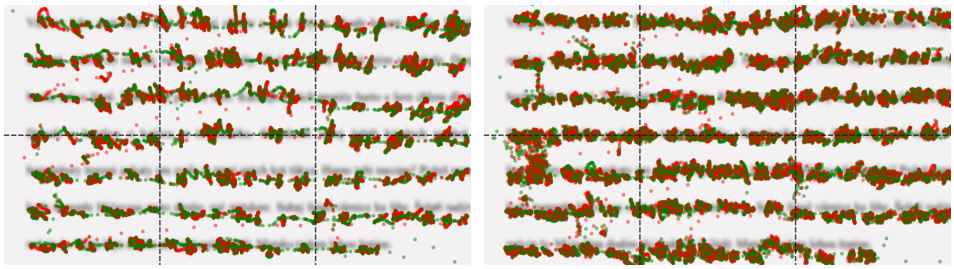
(**a**) Non-dyslexic (**b**) Dyslexic

**Task 1**



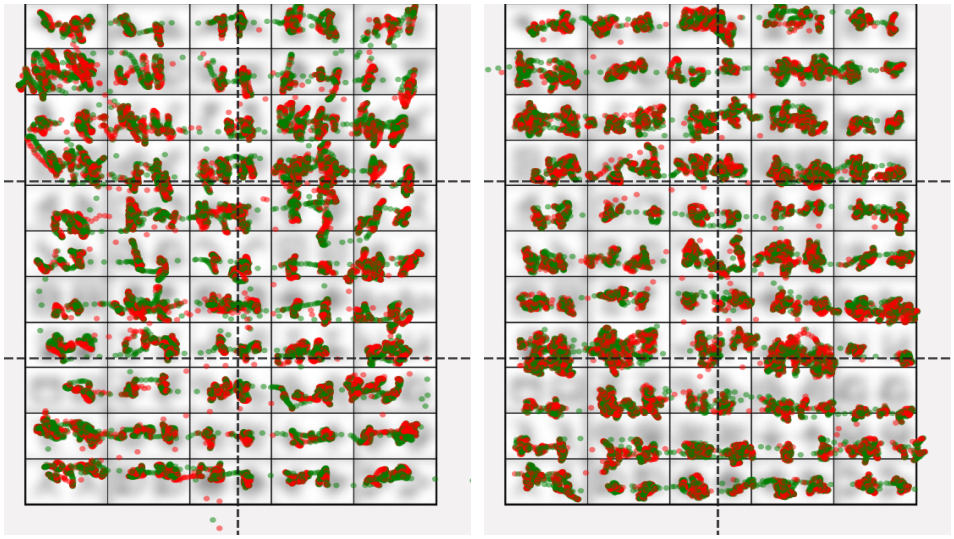
(**a**) Non-dyslexic (**b**) Dyslexic

**Task 2**



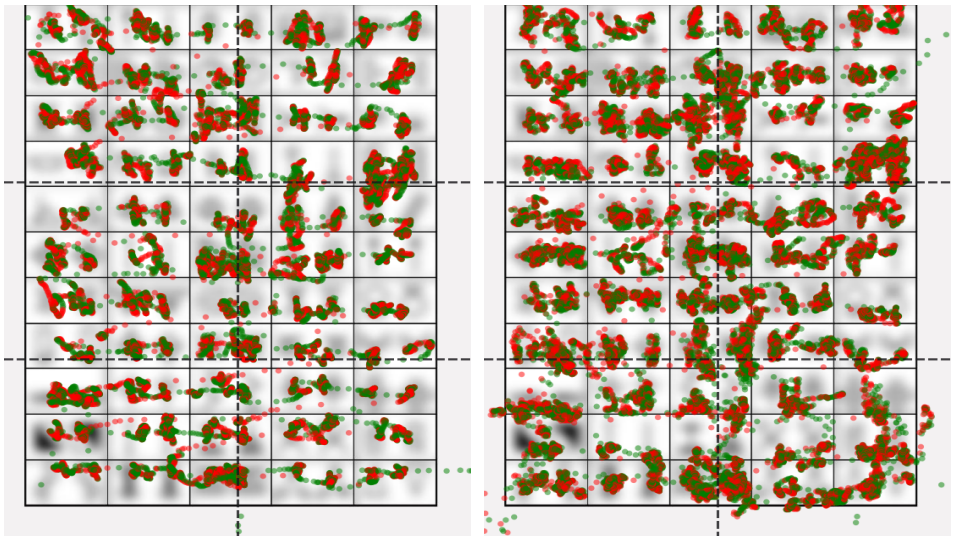
(**a**) Non-dyslexic (**b**) Dyslexic

**Task 3**



(**a**) Non-dyslexic (**b**) Dyslexic

**Task 4.1**



(**a**) Non-dyslexic (**b**) Dyslexic

**Task 4.2**

**Figure 4.3:** Comparison of Method RD-Sac on non-dyslexic and dyslexic participants on all tasks.

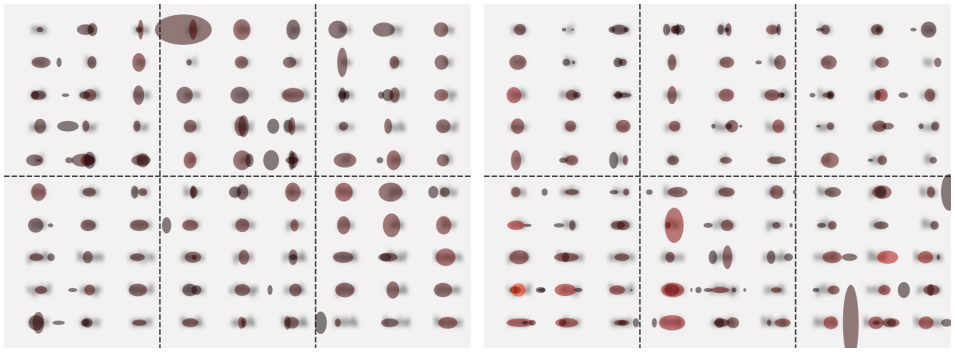
## 4.4 Fixation Image Approach

This section explains the fixation visualization method referenced in Section 2.3.2 and introduced in the recent studies [3, 16]. This thesis enhances this method by dividing the visualization image into sub-images and proposes alternative techniques to address issues arising from unwanted areas within these sub-images, as proposed in Section 4.4.1, Section 4.4.2, and Section 4.4.3.

### 4.4.1 Method Fix-D – Fixations with Duration Color

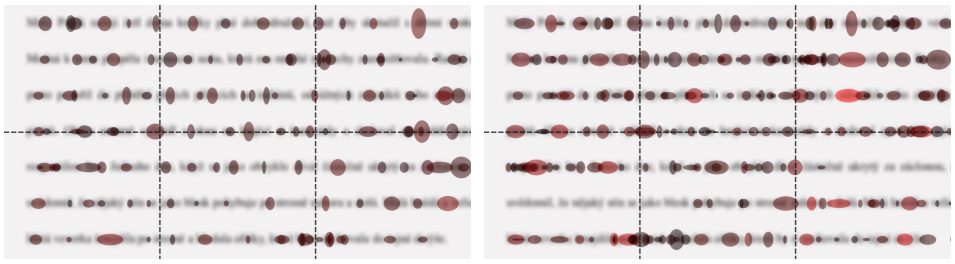
As mentioned, this method was introduced in those studies [3, 16]. This method visualizes *fixations* as ellipsoids instead of circles, where the size of the ellipsoid is determined by the dispersion of fixation in $x$ and $y$ directions. The ellipsoid's color represents the fixation duration, which is normalized for each task independently. Starting from black for shorter durations and transitioning through red to yellow for the longest durations. Figure 4.4 shows visualizations on all tested tasks

with comparison of dyslexic and non-dyslexic participants and with sub-image divided areas.



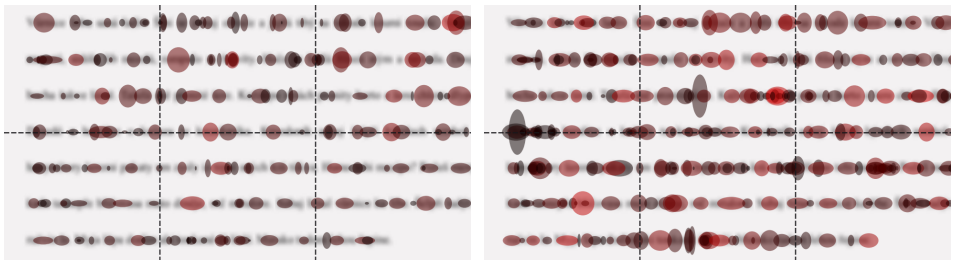
(**a**) Non-dyslexic (**b**) Dyslexic

**Task 1**



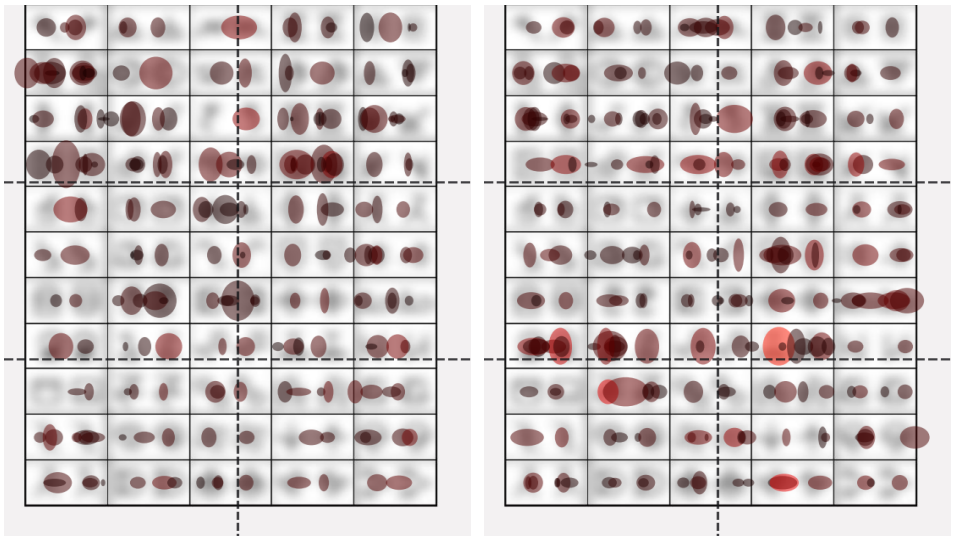
(**a**) Non-dyslexic (**b**) Dyslexic

**Task 2**



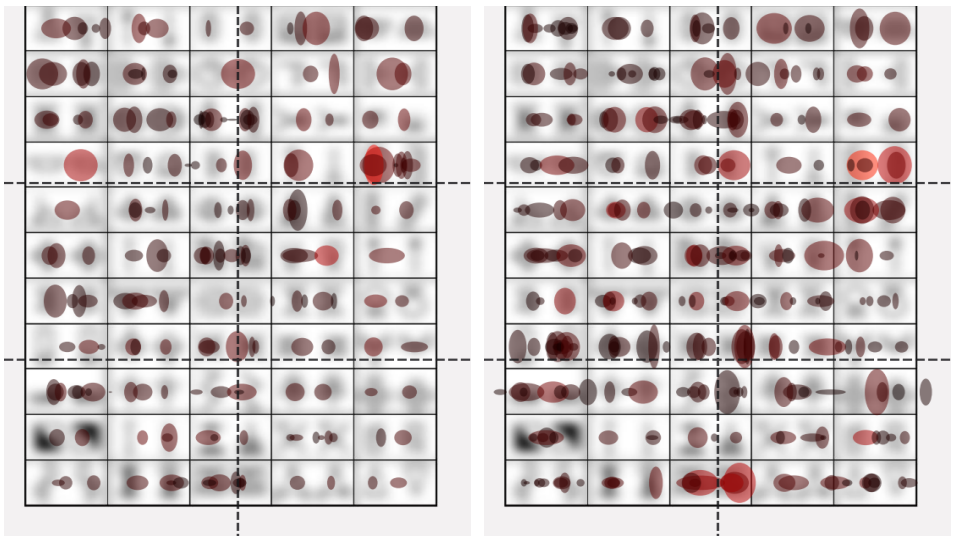
(**a**) Non-dyslexic (**b**) Dyslexic

**Task 3**



(**a**) Non-dyslexic (**b**) Dyslexic

**Task 4.1**



(**a**) Non-dyslexic (**b**) Dyslexic

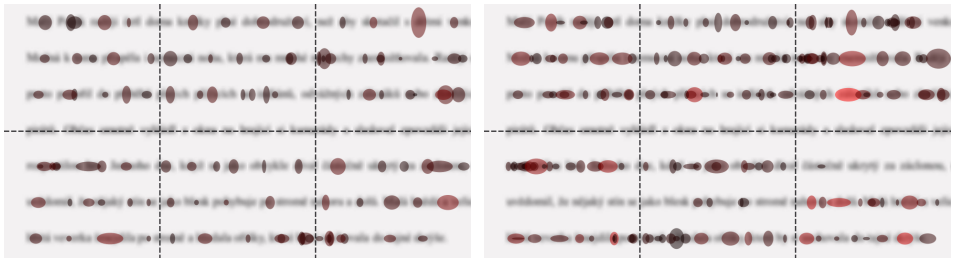
**Task 4.2**

**Figure 4.4:** Comparison of Method Fix-D on non-dyslexic and dyslexic participants on all tasks.

### 4.4.2 Method Fix-NO-D – No Overlapping Fixations with Duration Color

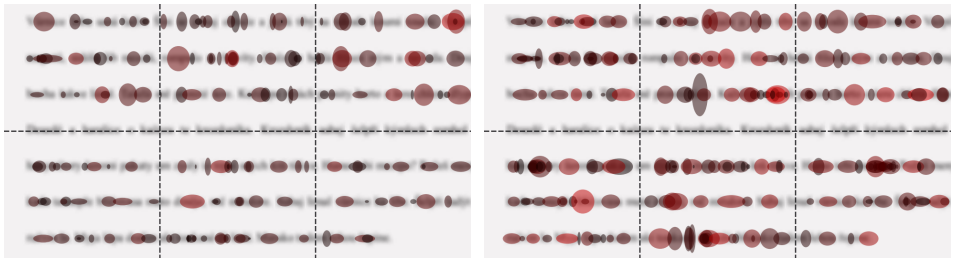
This method is based on visualizing fixations as ellipsoids with the size determined by the fixation dispersion and color by the normalized value of fixation duration as exactly described in Section 4.4.1. Since the image division process is based on the height and width of the fixed-size image, the division line went directly through some AOIs for some tasks. Especially for Task 2 and Task 3, the horizontal division line went through the whole sentence, as shown in Figure 4.5.

Fixations are considered overlapping when the height or width of a fixation ellipsoid extends from one sub-image into another. This method ignores all overlapping fixations. However, this leads to complete data loss of the middle sentence for Task 2 and Task 3. Other tasks are not shown in Figure 4.5 since there were only a few overlapping fixations.



(**a**) Non-dyslexic (**b**) Dyslexic

**Task 2**



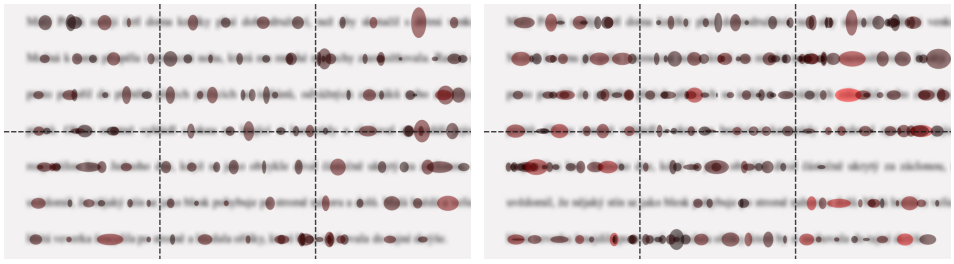
(**a**) Non-dyslexic (**b**) Dyslexic

**Task 3**

**Figure 4.5:** Comparison of Method Fix-NO-D on non-dyslexic and dyslexic participants on Task 2 and Task 3.

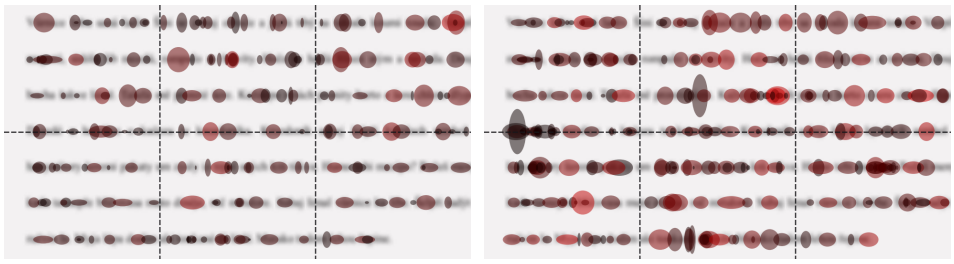
### 4.4.3 Method Fix-NHO-D – No Horizontal Overlapping Fixations with Duration Color

The Fix-NO-D method described in Section 4.4.2 resulted in the loss of the entire sentence. This method uses the exact visualization process as Fix-D method described in Section 4.4.1. Besides Fix-NO-D method in Section 4.4.2, this method keeps the overlapping fixations in the $y$ (vertical) direction and ignores those in the $x$ (horizontal) direction. Thus, the middle sentence in Task 2 and Task 3 is preserved as shown in Figure 4.6.



(**a**) Non-dyslexic (**b**) Dyslexic

**Task 2**



(**a**) Non-dyslexic (**b**) Dyslexic

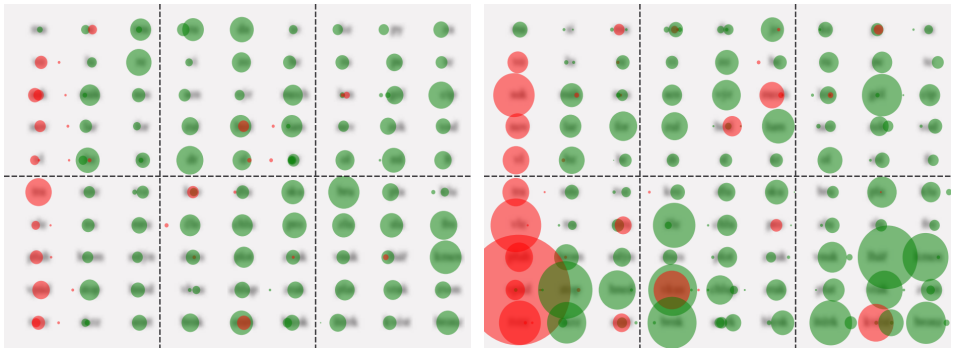
**Task 3**

**Figure 4.6:** Comparison of Method Fix-NHO-D on non-dyslexic and dyslexic participants on Task 2 and Task 3.

## 4.5 Method Fix-SD – Fixations with Saccade Color and Duration Size

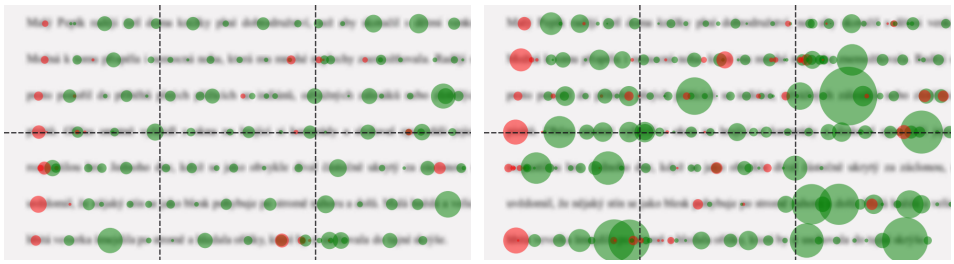
This method visualizes *fixations* as circles by utilizing forward and regression saccades as color. As in the RD-Sac method described in Section 4.3, green represents forward saccades, while red represents regression saccades. For the size (i.e., diameter) of the circle, the normalized value of fixation duration was utilized. Fixation durations were normalized based on the dataset's [3] minimal fixation duration value required for the I2MC algorithm to create fixations (40 ms) and on the highest fixation duration value for each task individually. The normalized duration value is then multiplied by a constant number 462 (i.e., parameter for tuning).

Utilizing fixation durations as circle size might also be beneficial, as dyslexic participants tend to have much longer fixation durations than non-dyslexic participants [25]. This observation can be seen in the images presented in Figure 4.7.



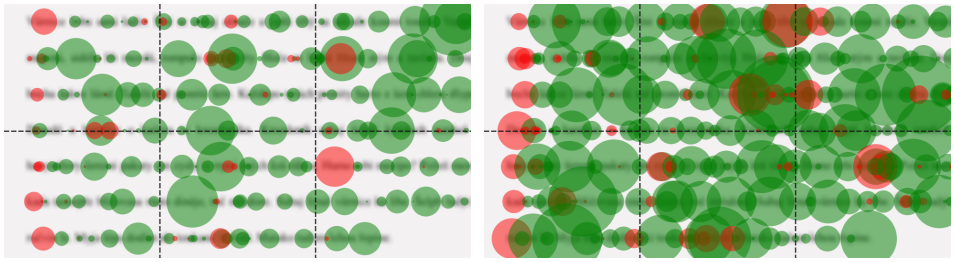
(**a**) Non-dyslexic (**b**) Dyslexic

**Task 1**



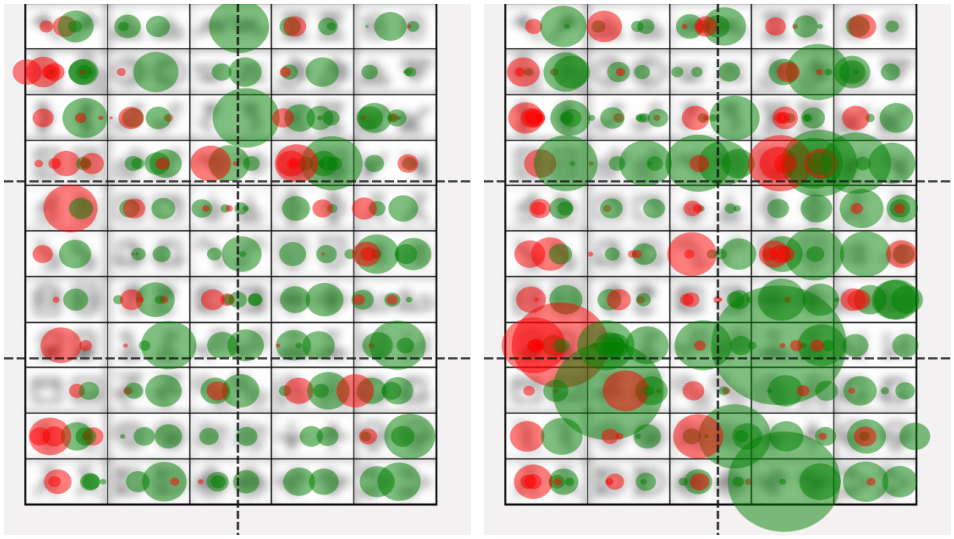
(**a**) Non-dyslexic (**b**) Dyslexic

**Task 2**



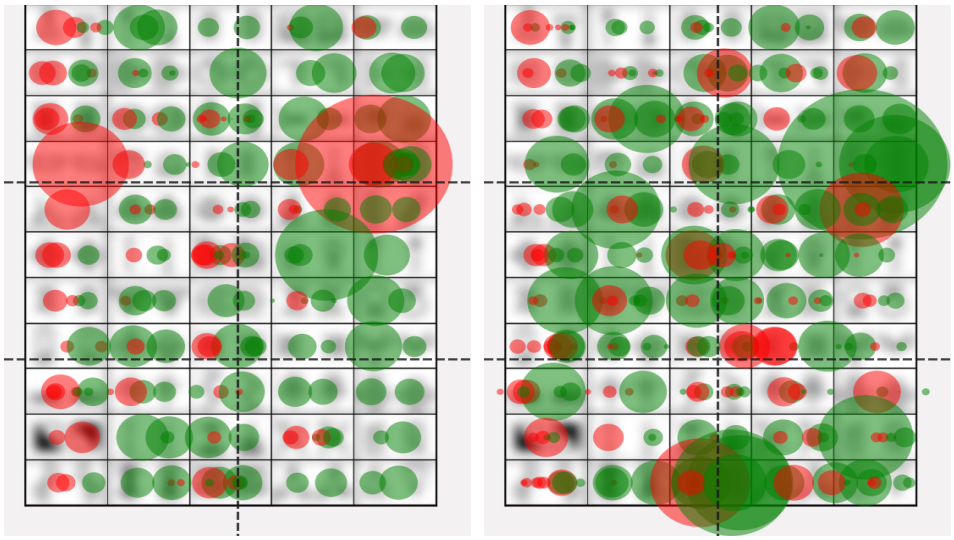
(**a**) Non-dyslexic (**b**) Dyslexic

**Task 3**



(**a**) Non-dyslexic (**b**) Dyslexic

**Task 4.1**



(**a**) Non-dyslexic (**b**) Dyslexic

**Task 4.2**

**Figure 4.7:** Comparison of Method Fix-SD on non-dyslexic and dyslexic participants on all tasks.

## 4.6 Classification and Majority Voting

For classification, this thesis uses a ResNet-18 model with pre-trained weights trained on general domain images with a fixed size of 224 × 224 pixels. Generated visualization sub-images are compressed into a fixed size of 224 × 224 pixels and serve as input to a ResNet-18 model. Those sub-images for each task and method are denoted by $I_1, \ldots, I_6$. Then, for each sub-image, the independent model is fine-tuned, creating six independent classifiers denoted as $cl_1(I_1), \ldots, cl_6(I_6)$ for each task and method. The classifier outputs a probability value $p \in [0, 1]$ denoted as $cl_i(I_i) = p_i$, representing the estimated likelihood that a participant exhibits characteristics associated with dyslexia. The output class is generated by a *bincls* function, indicating $1 = dyslexic$ if the probability is greater than or equal to a 50 % threshold, and $0 = non\text{-}dyslexic$ otherwise [4].

$$bincls(cl_i, I_i) = \begin{cases} 1 & \text{if } cl_i(I_i) \geq 0.5 \\ 0 & \text{else} \end{cases} \tag{4.1}$$

Using six classifiers for each method and task might seem excessive. However, combining their outputs into a single final decision for each task can simplify interpretation and enhance robustness. Some sub-images may lack informative features needed to distinguish between dyslexic and non-dyslexic participants, which can result in misleading classifier outputs due to the noise. To address this, two variants of majority voting were utilized to ensemble the outputs of the sub-image classifiers.

### 4.6.1 Binary Majority Voting

One of the simplest ensemble approaches is majority voting [26]. This thesis utilizes a binary version of majority voting, denoted as $mv$, since only two classes are required. The function $mv$ takes as input a set of $n$ classifiers and a set of their $n$ corresponding sub-images, where $n \in \{1, \ldots, 36\}$. The function outputs class 1 (dyslexic) if at least half of the classifiers, when applied to their respective sub-images, output class 1 (dyslexic). In case of even $n$, where there is exactly half of both classes, the function outputs class 1, since from the medical perspective, it might be preferable to minimize false negatives. Otherwise, it outputs class 0 (non-dyslexic).

$$mv(\{cl_1, \ldots, cl_n\}, \{I_1, \ldots, I_n\}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} bincls(cl_i, I_i) \geq \frac{n}{2} \\ 0 & \text{else} \end{cases} \tag{4.2}$$

This version of majority voting ignores the probability value $p$ and focuses only on the class. This leads to situations where the model can output a probability around 0.5, meaning that it is not sure about what class to put the input in. An example can be three classifiers $cl_1(I_1) = 0.54$, $cl_2(I_2) = 0.53$, $cl_3(I_3) = 0.06$ resulting in class 1 (dyslexic) for function $mv$. Since the results for $cl_1$ and $cl_2$ are almost random, $cl_3$ should be the indicator for the resulting class from the majority voting function. Similiary also for $cl_1(I_1) = 0.96, cl_2(I_2) = 0.42, cl_3(I_3) = 0.48$

resulting in class 0 (non-dyslexic) for function $mv$, however $cl_1$ should be the indicator for the resulting class.

### 4.6.2 Improved Binary Majority Voting

To address the problem with majority voting described in the previous Section 4.6.1, this version of binary majority voting, denoted as $imv$, focuses on probabilities $p$ instead of classes. The function $imv$ takes as input a set of $n$ classifiers and a set of their $n$ corresponding sub-images, where $n \in \{1, \ldots, 36\}$. The function outputs class 1 (dyslexic) if the sum of probabilities is greater than or equal to half of the total number of classifiers 1 (dyslexic). Otherwise, it outputs class 0 (non-dyslexic).

$$imv(\{cl_1, \ldots, cl_n\}, \{I_1, \ldots, I_n\}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} cl_i(I_i) \geq \frac{n}{2} \\ 0 & \text{else} \end{cases} \tag{4.3}$$

### 4.6.3 Majority Voting across Different Sub-Images

The six classifiers ($n = 6$) for each method and task can be ensembled together, resulting in one class output for each method and task. However, some sub-images may not contain the informative features necessary to differentiate between dyslexic and non-dyslexic participants. Using these sub-images may lead to misleading classifier outputs due to the noise.

Therefore, it may be necessary to disregard certain sub-images during the ensembling process ($n \in \{1, \ldots, 6\}$). Typically, those sub-images can be found by observing their classifier accuracy, where those with low accuracy are disregarded. This process might be time-consuming and handy. To improve efficiency, utilizing a brute-force method to find the best combination of sub-images with the most promising results for each method and task may be more effective.

### 4.6.4 Majority Voting across Different Methods

Some sub-images for different methods might provide different results, since some methods use different features for data visualization (e.g., raw eye-tracking data visualization and fixation data visualization). This might provide more stable results. Sub-images from

different methods can be ensembled together for each task. Thus, $n \in \{1, \ldots, 36\}$ denotes the possible ensemble of up to all six sub-images for each of the six methods, totaling 36 sub-images. This is done by using the best-performing combination of sub-images for each method described in the previous Section 4.6.3 and finding the best combination across different methods for each task. Due to the efficiency, this is also done by utilizing the brute-force method.

## 4.7 Implementation Details

For this thesis, an existing framework was used for experimental evaluation. This framework includes modules for feature extraction, a visualization generator, model training, and a final classification for new data. This framework is publicly available at: `https://gitlab.fi.muni.cz/xsedmid/dyslex`

This thesis extends the following modules of the framework:

- **Visualization generator** that generates newly introduced visualizations and divides them into sub-images. Each method described in this chapter has its own function named with the prefix `method_` and the method's name (e.g., `method_rd_idx` for method RD-Idx in Section 4.2). All functions are situated in file `visualization_generator.py`.

- **Majority voting** is for ensembling results from ResNets. Each version of majority voting described in Section 4.6 has its own function in file `majority_voting.py`. For binary majority voting, there is a function called `majority_vote`, and for the improved binary majority voting, a function `improved_majority_vote`.

The whole process for generating visualizations, model training, and the ensemble of results is summarized in a Jupyter notebook file `model_visualization_training.ipynb`.

Several methods offer a parameter *circle-size* that can be tuned to increase the model's performance. For this thesis, the well-performing values were found. For method RD-Idx, the circle size is set to 1.86 pixels, for RD-Sac it is 9.28 pixels, and for Fix-SD it is calculated as *normalized_duration* $\times$ 462 pixels.

# 5 Experimental Evaluation

This chapter outlines the experimental setup and evaluation procedures used to assess the classification performance of ResNet-18 models based on the proposed visualization methods. The aim is to determine how effectively individual ResNet-18 models can distinguish between dyslexic and non-dyslexic participants on sub-image inputs. Also, how the application of two different majority voting strategies can further enhance the overall classification accuracy by ensembling outputs from those models.

## 5.1 Methodology

For each task, a 5-fold cross-validation procedure was applied, resulting in five separate experiments. The final results were averaged across all experiments. The dataset of 70 participants was divided into folds with an equal number of dyslexic and non-dyslexic participants, using four folds for training and one for validation. The quality of the ResNet-18 classifier was determined by comparing the ratio of correctly classified participants from the validation fold [3]. The quality of majority voting functions was determined by the ratio of correctly classified participants to the total number of participants. The models were trained with a fixed learning rate ($lr = 0.0001$) and with 50 epochs.

## 5.2 Results of Proposed Visualization Methods

This section presents the classification performance of ResNet-18 models fine-tuned on all sub-images for each method and task. The evaluation focuses on how different visualization methods affect the model's ability to distinguish between dyslexic and non-dyslexic participants for each sub-image. The results emphasize the strengths of each method and help identify which visual feature encodings provide the most discriminative information for the specific task.

By looking at Table 5.1, some sub-images tend to have lower accuracy than others. The most effective method among all tested tasks

and methods is the newly introduced fixation-based visualization Fix-SD, which achieved an accuracy of up to 90 % for reading tasks and up to 78.57 % for visual-differentiation tasks. Among the best-scoring tasks was Task 3 (pseudo-text) with up to 90 % accuracy on some sub-images.

From raw eye-tracking visualizations, the more accurate method is RD-Idx with smaller circles and color based on timestamp order indexes than RD-Sac with bigger circles and color based on saccades. This is more likely due to the larger circles for RD-Sac, which resulted in too many circles overlapping other circles and also probably overshadowing possible patterns. Some RD-Idx sub-images, primarily for reading tasks, have accuracy similar to that of Fix-D method, which visualizes fixations. For the visual-differentiation tasks, one sub-image achieved an accuracy of 80 %. However, most of them achieved an accuracy below 70 %.

Fix-D, Fix-NO-D, and Fix-NHO-D performed similarly. Fix-NO-D and Fix-NHO-D methods offer a few sub-images with better accuracy than Fix-D. However, on some sub-images have lower accuracy than Fix-D. Surprisingly, Fix-NO-D method performed very well despite missing data from the whole sentence for Task 2 and Task 3. Compared with Fix-D, Fix-SD method offers better results for all tasks, primarily for Task 1 and Task 2 by achieving an accuracy of up to 88.57 % and for Task 3 by achieving an accuracy of 90 % on three sub-images and 88.57 % on other three sub-images. This is likely due to the increased occurrence of larger circles for dyslexic participants in Fix-SD method, as well as a higher number of regression saccades among dyslexic participants.

**Table 5.1:** Comparison of classification accuracies and standard deviations across all tested methods and tasks (the best results are highlighted in bold, the second best are underlined).

| Method | Sub-image | Task 1 | Task 2 | Task 3 | Task 4.1 | Task 4.2 |
|---|---|---|---|---|---|---|
| RD-Idx | 1 | 75.71 % ± 7.28 % | 84.29 % ± 5.35 % | 84.29 % ± 5.35 % | 61.43 % ± 7.28 % | 67.14 % ± 5.71 % |
| | 2 | 74.29 % ± 7.28 % | 84.29 % ± 8.33 % | 85.71 % ± 6.39 % | 58.57 % ± 7.00 % | 62.86 % ± 9.48 % |
| | 3 | 77.14 % ± 2.86 % | 82.86 % ± 5.71 % | 87.14 % ± 9.48 % | 68.57 % ± 9.69 % | 64.29 % ± 10.10 % |
| | 4 | 77.14 % ± 7.00 % | 85.71 % ± 4.52 % | 88.57 % ± 7.28 % | 61.43 % ± 5.71 % | 68.57 % ± 7.28 % |
| | 5 | 87.14 % ± 8.33 % | 84.29 % ± 8.33 % | 88.57 % ± 7.28 % | 67.14 % ± 7.28 % | 70.00 % ± 8.33 % |
| | 6 | 81.43 % ± 8.57 % | 82.86 % ± 5.71 % | 85.71 % ± 4.52 % | 70.00 % ± 8.33 % | **80.00** % ± 10.50 % |
| RD-Sac | 1 | 74.29 % ± 7.28 % | 78.57 % ± 10.10 % | 82.86 % ± 3.50 % | 65.71 % ± 10.50 % | 65.71 % ± 5.35 % |
| | 2 | 72.86 % ± 5.35 % | 77.14 % ± 5.35 % | 77.14 % ± 12.29 % | 57.14 % ± 7.82 % | 60.00 % ± 7.28 % |
| | 3 | 77.14 % ± 5.35 % | 78.57 % ± 4.52 % | 80.00 % ± 2.86 % | 67.14 % ± 9.69 % | 68.57 % ± 8.57 % |
| | 4 | 72.86 % ± 8.33 % | 74.29 % ± 7.28 % | 78.57 % ± 11.95 % | 68.57 % ± 3.50 % | 65.71 % ± 8.33 % |
| | 5 | 75.71 % ± 7.28 % | 80.00 % ± 8.33 % | 77.14 % ± 13.85 % | 58.57 % ± 10.50 % | 75.71 % ± 9.69 % |
| | 6 | 80.00 % ± 10.50 % | 81.43 % ± 3.50 % | 72.86 % ± 8.33 % | 64.29 % ± 9.04 % | 61.43 % ± 8.57 % |
| Fix-D | 1 | 78.57 % ± 6.39 % | 82.86 % ± 7.28 % | **90.00** % ± 7.28 % | 62.86 % ± 8.33 % | 62.86 % ± 10.50 % |
| | 2 | 80.00 % ± 9.48 % | 81.43 % ± 5.71 % | 88.57 % ± 9.69 % | 65.71 % ± 8.33 % | 61.43 % ± 7.28 % |
| | 3 | 81.43 % ± 5.71 % | 82.86 % ± 5.71 % | 84.29 % ± 8.33 % | 62.86 % ± 8.33 % | 62.86 % ± 2.86 % |
| | 4 | 77.14 % ± 5.35 % | 81.43 % ± 7.28 % | 87.14 % ± 5.35 % | 60.00 % ± 10.69 % | 72.86 % ± 7.00 % |
| | 5 | 80.00 % ± 8.33 % | 85.71 % ± 7.82 % | 85.71 % ± 7.82 % | 57.14 % ± 4.52 % | 64.29 % ± 6.39 % |
| | 6 | 82.86 % ± 11.61 % | 84.29 % ± 5.35 % | 82.86 % ± 7.28 % | 65.71 % ± 12.29 % | 70.00 % ± 2.86 % |
| Fix-NO-D | 1 | 81.43 % ± 5.71 % | 87.14 % ± 7.00 % | 85.71 % ± 6.39 % | 57.14 % ± 4.52 % | 65.71 % ± 10.50 % |
| | 2 | 78.57 % ± 11.95 % | 84.29 % ± 5.35 % | 85.71 % ± 7.82 % | 65.71 % ± 8.33 % | 61.43 % ± 7.28 % |
| | 3 | 82.86 % ± 7.28 % | 85.71 % ± 4.52 % | 88.57 % ± 7.28 % | 67.14 % ± 8.57 % | 58.57 % ± 5.35 % |
| | 4 | 81.43 % ± 3.50 % | 78.57 % ± 6.39 % | 82.86 % ± 3.50 % | 57.14 % ± 9.04 % | 65.71 % ± 12.29 % |
| | 5 | 78.57 % ± 7.82 % | 80.00 % ± 8.33 % | 88.57 % ± 5.71 % | 62.86 % ± 5.35 % | 62.86 % ± 5.35 % |
| | 6 | 87.14 % ± 9.48 % | 81.43 % ± 5.71 % | 85.71 % ± 7.82 % | 61.43 % ± 7.28 % | 65.71 % ± 9.48 % |
| Fix-NHO-D | 1 | 81.43 % ± 9.69 % | **88.57** % ± 7.28 % | 88.57 % ± 7.28 % | 57.14 % ± 4.52 % | 65.71 % ± 8.33 % |
| | 2 | 78.57 % ± 10.10 % | 81.43 % ± 3.50 % | 87.14 % ± 5.35 % | 64.29 % ± 4.52 % | 57.14 % ± 7.82 % |
| | 3 | 82.86 % ± 7.28 % | 85.71 % ± 7.82 % | 87.14 % ± 7.00 % | 61.43 % ± 9.69 % | 61.43 % ± 5.71 % |
| | 4 | 81.43 % ± 3.50 % | 82.86 % ± 7.28 % | 85.71 % ± 6.39 % | 55.71 % ± 8.33 % | 68.57 % ± 13.25 % |
| | 5 | 80.00 % ± 7.00 % | 87.14 % ± 7.00 % | 87.14 % ± 5.35 % | 60.00 % ± 7.28 % | 61.43 % ± 7.28 % |
| | 6 | 84.29 % ± 8.33 % | 84.29 % ± 2.86 % | 82.86 % ± 5.71 % | 62.86 % ± 8.33 % | 68.57 % ± 7.28 % |
| Fix-SD | 1 | 78.57 % ± 7.82 % | **88.57** % ± 9.69 % | 88.57 % ± 9.69 % | 71.43 % ± 10.10 % | 70.00 % ± 8.33 % |
| | 2 | **88.57** % ± 7.28 % | 82.86 % ± 9.69 % | 88.57 % ± 9.69 % | 64.29 % ± 9.04 % | 65.71 % ± 11.43 % |
| | 3 | **88.57** % ± 11.61 % | 85.71 % ± 7.82 % | **90.00** % ± 7.28 % | **75.71** % ± 9.69 % | 71.43 % ± 10.10 % |
| | 4 | 87.14 % ± 9.48 % | **88.57** % ± 7.28 % | **90.00** % ± 7.28 % | 70.00 % ± 5.35 % | 78.57 % ± 11.95 % |
| | 5 | 85.71 % ± 6.39 % | 84.29 % ± 13.85 % | 88.57 % ± 8.57 % | 68.57 % ± 9.69 % | 71.43 % ± 14.29 % |
| | 6 | **88.57** % ± 11.61 % | 85.71 % ± 10.10 % | **90.00** % ± 8.57 % | 62.86 % ± 5.35 % | 72.86 % ± 8.33 % |

## 5.3 Results of Majority Voting across Different Sub-Images

Results in Table 5.2 were obtained by ensembling all sub-images for each method and each task. The ensemble results were slightly better than those described in the previous Section 5.2. Some methods, primarily on reading tasks, exceed an accuracy of 90 %. The difference between normal (i.e., based on the dyslexic and non-dyslexic classes) and improved (i.e., based on the classifier's output probability) majority voting for reading tasks is up to 10 percentage points. For the visual-differentiation tasks, the difference is up to 16 percentage points. In general, the normal version of majority voting provides better accu-

racy. The improved version of majority voting results in lower accuracy. It is important to note that the ResNet-18 models were fine-tuned only on the dataset of 70 participants, so their output probabilities may not be very reliable.

Among the results based on the raw eye-tracking data, RD-Idx method outperforms RD-Sac method and is comparable to fixation methods at the same level by achieving an accuracy of up to 90 % for reading tasks. Notably, RD-Sac method achieved higher accuracy on visual-differentiation tasks with normal majority voting than RD-Idx method. However, the difference between normal and improved majority voting is up to 15.72 percentage points, much higher than RD-Idx difference, which is up to 7.14 percentage points.

Among the best fixation visualization methods, Fix-SD shows significant improvement, primarily in Task 1 and visual-differentiation tasks. For Task 2, other fixation visualization methods outperform Fix-SD on normal majority voting by achieving higher accuracy of up to 91.43 %. In the case of improved majority voting, Fix-SD still ranks among the best ones.

**Table 5.2:** Comparison of majority voting results on all tasks for each method, including all six sub-images (the best results are highlighted in bold, the second best are underlined).

| Method | Type | Task 1 | Task 2 | Task 3 | Task 4.1 | Task 4.2 |
|---|---|---|---|---|---|---|
| RD-Idx | Normal | 84.29 % | 85.71 % | <u>90.00</u> % | 68.57 % | 77.14 % |
| | Improved | 77.14 % | 84.29 % | 87.14 % | 64.29 % | 70.00 % |
| RD-Sac | Normal | 82.86 % | 82.86 % | 82.86 % | <u>72.86</u> % | <u>78.57</u> % |
| | Improved | 72.86 % | 80.00 % | 72.86 % | 57.14 % | 65.71 % |
| Fix-D | Normal | 90.00 % | <u>88.57</u> % | **91.43** % | 61.43 % | 71.43 % |
| | Improved | 81.43 % | 82.86 % | 84.29 % | 60.00 % | 57.14 % |
| Fix-NO-D | Normal | <u>91.43</u> % | <u>88.57</u> % | 88.57 % | 61.43 % | 64.29 % |
| | Improved | 84.29 % | 80.00 % | 88.57 % | 55.71 % | 55.71 % |
| Fix-NHO-D | Normal | 88.57 % | **90.00** % | 88.57 % | 60.00 % | 62.86 % |
| | Improved | 81.43 % | 82.86 % | 84.29 % | 48.57 % | 55.71 % |
| Fix-SD | Normal | **92.86** % | 87.14 % | **91.43** % | **77.14** % | **81.43** % |
| | Improved | 84.29 % | 85.71 % | 88.57 % | 67.14 % | 74.29 % |

In comparison with the binary classification implemented by Sedmidubsky et al. [3], the ensembled Fix-D method provides better

results, particularly on normal majority voting, for all tasks. For Task 1 (90.00 % on normal and 81.43 % on improved majority voting vs. 82.86 %), for Task 2 (88.57 % on normal and 82.86 % on improved majority voting vs. 82.86 %), and for Task 3 (91.43 % on normal and 84.29 % on improved majority voting vs. 88.57 %).

All possible combinations of sub-images were evaluated, and the results presented in Table 5.3 show the ensemble of the best-performing combination of sub-images for each method and task. Notably, the results are more consistent, indicating that there is a lower difference between normal and improved majority voting. With this approach, many methods achieved over 90 % accuracy on reading tasks. Primarily for Task 3, where Fix-D and Fix-SD achieved an accuracy of 92.86 % on normal majority voting and 91.43 % on improved. Notably on Task 3, Fix-NHO-D achieved an accuracy of 94.29 % on normal majority voting, while 87.14 % on improved. For visual-differentiation tasks, RD-Idx and Fix-SD achieved an accuracy of up to 82.86 % on normal majority voting.

For an individual task, the sub-images vary for an individual method (e.g., Task 1 has 5 and 6 sub-images for method RD-Idx, while method RD-Sac has 1, 3, and 6 sub-images). The variability is likely due to the features differently encoded (e.g., raw eye-tracking data vs. fixations, different sizes of circles or ellipsoids) by each visualization method. This can impact how effectively the ResNet-18 model learns meaningful patterns. Consequently, some sub-images may highlight more discriminative information, while others might introduce noise or less relevant structures.

**Table 5.3:** Comparison of majority voting results for all tested methods, tasks, and sub-image combinations providing the best performance (the best results are highlighted in bold, the second best are underlined).

| Method | Type | Task 1 | Task 2 | Task 3 | Task 4.1 | Task 4.2 |
|---|---|---|---|---|---|---|
| RD-Idx | Sub-images | 5, 6 | 1, 2, 5 | 2, 3, 4, 5, 6 | 2, 3, 4, 5, 6 | 4, 5, 6 |
| | Normal | 85.71 % | 88.57 % | 91.43 % | **82.86** % | <u>80.00</u> % |
| | Improved | 85.71 % | 87.14 % | 87.14 % | 65.71 % | 74.29 % |
| RD-Sac | Sub-images | 1, 3, 6 | 1, 3, 6 | 1, 3 | 1, 2, 3, 4, 6 | 3, 4, 5 |
| | Normal | 81.43 % | 85.71 % | 82.86 % | 75.71 % | <u>80.00</u> % |
| | Improved | 80.00 % | 84.29 % | 81.43 % | 62.86 % | 71.43 % |
| Fix-D | Sub-images | 1, 4, 5, 6 | 1, 3, 4, 6 | 1, 2, 3, 4 | 2, 3, 4, 5, 6 | 2, 4, 6 |
| | Normal | 90.00 % | 90.00 % | <u>92.86</u> % | 67.14 % | 77.14 % |
| | Improved | 84.29 % | 85.71 % | 91.43 % | 62.86 % | 71.43 % |
| Fix-NO-D | Sub-images | 1, 3, 6 | 1, 2, 3, 5, 6 | 1, 2, 3, 4, 5 | 1, 3, 6 | 1, 4, 5 |
| | Normal | **92.86** % | 90.00 % | 90.00 % | 70.00 % | 72.86 % |
| | Improved | 84.29 % | 84.29 % | 88.57 % | 64.29 % | 60.00 % |
| Fix-NHO-D | Sub-images | 1, 4, 6 | 1, 3, 6 | 1, 3, 4 | 2, 3, 5 | 1, 4, 6 |
| | Normal | <u>91.43</u> % | <u>91.43</u> % | **94.29** % | 64.29 % | 70.00 % |
| | Improved | 85.71 % | 88.57 % | 87.14 % | 58.57 % | 65.71 % |
| Fix-SD | Sub-images | 4, 6 | 1, 3, 5, 6 | 3, 4, 5, 6 | 1, 2, 3 | 1, 2, 4, 5 |
| | Normal | **92.86** % | **92.86** % | <u>92.86</u> % | <u>80.00</u> % | **82.86** % |
| | Improved | 88.57 % | 88.57 % | 91.43 % | 72.86 % | 77.14 % |

## 5.4 Results of Majority Voting across Different Methods

Results in Table 5.4 combine different methods and their best combination of sub-images from Table 5.3. All reading tasks achieved an accuracy of 92.86 % for normal majority voting, and the maximum difference between normal and improved majority voting is 4.29 percentage points for Task 2. For Task 3, both Fix-D and Fix-SD independently achieved the same accuracy of 92.86 % on normal and 91.43 % on improved majority voting. Visual-differentiation tasks achieved an accuracy of up to 88.57 % on normal majority voting, while the maximum difference between normal and improved majority voting is 17.14 percentage points for Task 4.1.

**Table 5.4:** Comparison of majority voting results on all tasks across different methods using the best-performing combination of sub-images from Table 5.3.

| Task | Methods | Type | Score |
|---|---|---|---|
| Task 1 | Fix-NO-D, Fix-SD | Normal | 92.86 % |
| | | Improved | 90.00 % |
| Task 2 | Fix-SD | Normal | 92.86 % |
| | | Improved | 88.57 % |
| Task 3 | Fix-D or Fix-SD | Normal | 92.86 % |
| | | Improved | 91.43 % |
| Task 4.1 | RD-Idx, RD-Sac, Fix-NO-D, Fix-SD | Normal | 88.57 % |
| | | Improved | 71.43 % |
| Task 4.2 | RD-Idx, RD-Sac, Fix-SD | Normal | 87.14 % |
| | | Improved | 78.57 % |

## 5.5 Quantitative Results

The visualization generation process was executed on a machine equipped with an AMD Ryzen 5 2600 processor, while training of the ResNet-18 models was carried out on an NVIDIA GeForce GTX 1660 Ti graphics card. Generating a single sub-image took approximately 1.6 seconds for the raw eye-tracking data methods and around 0.4 seconds for the methods based on fixations. Training a ResNet-18 model on one sub-image required approximately 23 minutes.

# 6 Conclusions

This thesis aimed to improve the classification accuracy of state-of-the-art approaches using ResNet-18. This was done by developing newly generated visualizations containing raw eye-tracking data, fixations, and saccades. Due to the pre-trained weights of ResNet-18 causing the fixed size of the input image, the visualizations were divided into sub-images that were used as input. Results from each individual sub-image classifier were then ensembled using two different versions of majority voting. This process achieved an accuracy of up to 92.86 % for reading tasks and an accuracy of up to 88.57 % for newly introduced visual-differentiation tasks.

Visualizations based on raw eye-tracking data are generally not as effective as those based on fixations, but still provide promising results. When comparing the best-performing combinations of sub-images, the most effective method based on raw eye-tracking data performed only 1.43—7.15 percentage points worse in accuracy than the best method based on fixations. Among all tasks and tested methods, the most promising method was the newly introduced Fix-SD method.

The process of splitting entire visualization images into smaller sub-images is very promising because it allows the focus to be only on specific sub-images that may indicate significant differences between dyslexic and non-dyslexic participants. For comparison, the best sub-image for Fix-SD method achieved an accuracy of 88.57 %, while the worst achieved 78.57 % accuracy in Task 1. Although the image-splitting method was not ideal due to the division affecting some AOIs, the results were still promising.

For future research, selecting a more effective method for image division might be beneficial, and also by using more sub-images focusing on individual sentences with proportionally empty surroundings for reading tasks. For the grid form used in visual-differentiation tasks, having a sub-image for each pair of symbols might be useful, but this may be time-consuming with lots of pairs of symbols. Additionally, a grid layout with an even number of columns and rows, with larger gaps between pairs, might be more helpful for systematic image division. Also, encoding information about correct and incorrect clicks as color may be more informative than encoding saccades.

# Bibliography


1. EMILY M. LIVINGSTON, Linda S. Siegel; RIBARY, Urs. Developmental dyslexia: emotional impact and consequences. *Australian Journal of Learning Difficulties*. 2018, vol. 23, no. 2, pp. 107–135. Available from DOI: 10.1080/19404158.2018.1479975.
2. YANG, L.; LI, C.; LI, X.; ZHAI, M.; AN, Q.; ZHANG, Y.; ZHAO, J.; WENG, X. Prevalence of Developmental Dyslexia in Primary School Children: A Systematic Review and Meta-Analysis. *Brain Sciences*. 2022, vol. 12, no. 2, p. 240. Available from DOI: 10.3390/brainsci12020240.
3. SEDMIDUBSKY, Jan; DOSTALOVA, Nicol; SVARICEK, Roman; CULEMANN, Wolf. ETDD70: Eye-Tracking Dataset for Classification of Dyslexia Using AI-Based Methods. In: *Similarity Search and Applications*. Cham: Springer Nature Switzerland, 2025, pp. 34–48. ISBN 978-3-031-75823-2.
4. ŠVAŘÍČEK, Roman; DOSTÁLOVÁ, Nicol; SEDMIDUBSKÝ, Jan; ČERNEK, Andrej. INSIGHT: Combining Fixation Visualizations and Residual Neural Networks for Dyslexia Classification from Eye-Tracking Data. *DYSLEXIA*. 2025, vol. 31. ISSN 1076-9242. Available from DOI: http://dx.doi.org/10.1002/dys.1801.
5. HABIB, Michel. The neurological basis of developmental dyslexia: An overview and working hypothesis. *Brain*. 2000, vol. 123, no. 12, pp. 2373–2399. ISSN 0006-8950. Available from DOI: 10.1093/brain/123.12.2373.
6. SNOWLING, Margaret J. *Dyslexia: A Very Short Introduction*. Oxford, UK: Oxford University Press, 2013. ISBN 9780191663569.
7. NYSTRÖM, Marcus; HOOGE, Ignace T. C.; HESSELS, Roy S.; ANDERSSON, Richard; HANSEN, Dan Witzner; JOHANSSON, Roger; NIEHORSTER, Diederick C. The fundamentals of eye tracking part 3: How to choose an eye tracker. *Behavior Research Methods*. 2025, vol. 57, no. 67. Available from DOI: 10.3758/s13428-024-02587-x.

8. RAYNER, Keith. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*. 1998, vol. 124, no. 3, pp. 372–422.

9. BLASCHECK, Tanja; KURZHALS, Kuno; RASCHKE, Michael; BURCH, Michael; WEISKOPF, Daniel; ERTL, Thomas. State-of-the-art of visualization for eye tracking data. In: *Eurovis (stars)*. 2014, p. 29.

10. HE, Kaiming; ZHANG, Xiangyu; REN, Shaoqing; SUN, Jian. Deep Residual Learning for Image Recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.

11. KAISAR, Shahriar. Developmental dyslexia detection using machine learning techniques : A survey. *ICT Express*. 2020, vol. 6, no. 3, pp. 181–184. ISSN 2405-9595. Available from DOI: https://doi.org/10.1016/j.icte.2020.05.006.

12. RELLO, Luz; BALLESTEROS, Marcos. Detecting Readers with Dyslexia Using Machine Learning with Eye Tracking Measures. In: *Proceedings of the 12th Web for All Conference*. Association for Computing Machinery, 2015, 16:1–16:8. Available from DOI: 10.1145/2745555.2746644.

13. NILSSON BENFATTO, Mattias; ÖQVIST SEIMYR, Gustaf; YGGE, Jan; PANSELL, Tony; RYDBERG, Agneta; JACOBSON, Christer. Screening for Dyslexia Using Eye Tracking during Reading. *PLOS ONE*. 2016, vol. 11, no. 12, e0165508. Available from DOI: 10.1371/journal.pone.0165508.

14. NERUŠIL, Boris; POLEC, Jaroslav; ŠKUNDA, Juraj; KAČUR, Juraj. Eye tracking based dyslexia detection using a holistic approach. *Scientific Reports*. 2021, vol. 11, p. 15687. Available from DOI: 10.1038/s41598-021-95275-1.

15. VAJS, Ivan; KOVIĆ, Vanja; PAPIĆ, Tamara; SAVIĆ, Andrej M.; JANKOVIĆ, Milica M. Spatiotemporal Eye-Tracking Feature Set for Improved Recognition of Dyslexic Reading Patterns in Children. *Sensors*. 2022, vol. 22, no. 13. ISSN 1424-8220. Available from DOI: 10.3390/s22134900.

16. ČERNEK, Andrej. *Recognition of Reading Disorder Based on Eye-Tracking Data*. Brno, 2023. Available also from: https://is.muni.cz/th/ib5jl/. Diplomová práce. Masarykova univerzita, Fakulta informatiky. Supervised by Jan SEDMIDUBSKÝ.

17. LABORATORIUM ELEKTROTECHNIKI I ELEKTRONIKI OKRĘTOWEJ. *Urządzenia - Laboratorium Elektrotechniki i Elektroniki Okrętowej*. 2025. Available also from: http://www.lelo.uw.edu.pl/urzadzenia. Accessed: 2025-04-27.

18. DOVBNYCH, Maryna; PLECHAWSKA–WÓJCIK, Małgorzata. A comparison of conventional and deep learning methods of image classification. *Journal of Computer Sciences Institute*. 2021, vol. 21, pp. 303–308. Available from DOI: 10.35784/jcsi.2727.

19. WANG, Yuheng. Research on Image Classification Based on ResNet. In: *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)*. Atlantis Press, 2023, pp. 971–979. ISBN 978-94-6463-300-9. ISSN 2352-538X. Available from DOI: 10.2991/978-94-6463-300-9_98.

20. MITCHELL, Tom M. *Machine Learning*. New York: McGraw-Hill, 1997. Available also from: https://www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf.

21. NASTESKI, Vladimir. An overview of the supervised machine learning methods. *HORIZONS.B*. 2017, vol. 4, pp. 51–62. Available from DOI: 10.20544/HORIZONS.B.04.1.17.P05.

22. MURALI, Kaushik; SURANI, Isha; BHARDWAJ, Aviral; JAIN, Ananya. A Comparative Study of CNN, ResNet, and Vision Transformers for Multi-Classification of Chest Diseases. 2024. Available from arXiv: 2406.00237 [eess.IV].

23. BEDNÁŘOVÁ, J. *Diagnostika schopností a dovedností v oblasti čtení a psaní. Varianta pro pedagogy škol a školní poradenská pracoviště, 3.—4. ročník*. Pedagogickopsychologická poradna Brno, 2015.

24. HESSELS, R.S.; NIEHORSTER, D.C.; KEMNER, C., et al. Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). *Behavior Research Methods*. 2017, vol. 49, no. 5, pp. 1802–1823. Available from DOI: 10.3758/s13428-016-0822-1.

25. HUTZLER, Florian; KRONBICHLER, Martin; JACOBS, Arthur M.; WIMMER, Heinz. Perhaps Correlational but Not Causal: No Effect of Dyslexic Readers' Magnocellular System on Their Eye Movements During Reading. *Neuropsychologia*. 2006, vol. 44, no. 4, pp. 637–648. Available from DOI: 10.1016/j.neuropsychologia.2005.06.006.

26. DOGAN, Alican; BIRANT, Derya. A Weighted Majority Voting Ensemble Approach for Classification. In: *2019 4th International Conference on Computer Science and Engineering (UBMK)*. 2019, pp. 1–6. Available from DOI: 10.1109/UBMK.2019.8907028.

# A Content of the Electronic Appendix

Part of this thesis involves an extension of the used framework, which can be found in the archive named `FrameworkExtension.zip`. A summary of the extended modules is described in Section 4.7 and is also included in the README file located in the attached archive.