Deep Learning Models to Meet Eye Fixation and Dyslexia

Maria Krylova
Research assistant at AICS
Dr. Soroosh Shalileh
Head of AICS and Research fellow at CLB

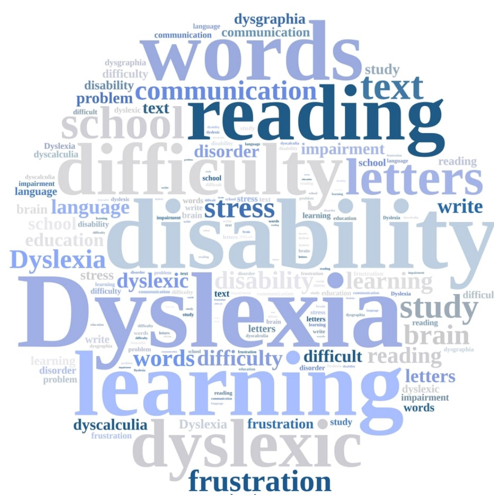
Regular Scientific Seminars of Laboratory of Artificial Intelligence for Cognitive Sciences (AICS), HSE University
15 May 2024, Moscow, Russia.

Contents

Introduction

Problem

- **Developmental dyslexia:** a learning disorder characterized by specific reading impairment despite normal intelligence and oral language skills *Frazier, M. 2016.*
- **According to** www.dyslexia-reading-well.com **approximately 15% of people have dyslexia.**
- **This significant percentage of our population deserves some cares**



https://dma.org.uk/article/my-dyslexia-journey

NATIONAL RESEARCH UNIVERSITY

Introduction

motivation and novelty

- **Goal:** find the most effective "data representation" for eye-tracking data to develop a robust artificial intelligence system to identify dyslexia at the early stages
- **Novelty:** focusing solely on eye-fixation data for detecting dyslexia.
- This research answers the following questions:
    - Q1: Is it possible to predict dyslexia only from the eye fixation data?
    - Q2: How informative is eye fixation data in predicting absence or presence of dyslexia?
    - Q3: How can eye fixation data be represented for training AI models in predicting dyslexia?
    - Q4: What family of Artificial Intelligence (AI) methods is the most effective at predicting dyslexia?

NATIONAL RESEARCH UNIVERSITY

Laboratory of Artificial Intelligence for Cognitive Sciences

Literature review

Previous methods

- According to [1], the latest and most comprehensive review on the applications of AI to identify dyslexia:
    1. Support vector machines, multi-layer perceptron, and random forest, in descending order, are the three most commonly applied AI classification algorithms.
    2. Majority of the previous research obtained approximately, in the best case, scenario, 85% – 88 % accuracy (ROC AUC)

[1]: Usman OL, Muniyandi RC, Omar K, Mohamad M. Advance machine learning methods for Dyslexia bio- marker detection: a review of implementation details and challenges. IEEE Access. 2021; 9:36879– 36897.

Literature review

Previous methods

3. Seven types of data (MRI, fMRI, face video/image, reading test errors, test scores, EEG, and eye tracking) are used to train AI models.
4. Considering the number of unique data sets:

   eye-tracking-based (seven data sets), EEG (six data sets), and MRI (five data sets) are the top three frequently used data types.

- Eye tracking are directly related to reading process and are not intrusive.

Literature review

Previous methods

- Refer to:
  - [2] for more information on the latest technological developments of AI and eye-fixation;
  - [3] for review of Eye-tracking Technology in Dyslexia Diagnosis
  - [4] for a short review of on this subject in Russian Language

[2] Shalileh S, Ignatov D, Lopukhina A, Dragoy O. Identifying dyslexia in school pupils from eye movement and demographic data using artificial intelligence. Plos one. 2023 Nov 22;18(11):e0292047.

[3] Coenen L, Grünke M, Becker-Genschow S, Schlüter K, Schulden M, Barwasser A. A Systematic Review of Eye-Tracking Technology in Dyslexia Diagnosis. Insights into Learning Disabilities. May 2024 21(1), 45-65.

[4] Грачева МА, Шалилех С. Диагностика дислексии с использованием методов искусственного интеллекта по данным движений глаз: обзор. Клиническая и специальная психология. 2023;12(3):1-29.

NATIONAL RESEARCH UNIVERSITY

AICS
Laboratory of Artificial Intelligence for Cognitive Sciences

Literature review

Eye Fixation data sets

Control Group Size	High Risk of Dyslexia Size	Low Risk of Dyslexia Size	Age Range	Target Values		Language
				Discrete	Continuous	
97	88	0	9-10	2	-	Swedish
32	37	0	8.5-12.5	2	-	Greek
18	18	0	8-12	2	2	Serbian
135	30	0	ave. 12.5	2	-	Finnish
41	46	0	12.3-18	2	-	French
49	48	0	11-55	2	-	Spanish
213	72	22	6-14	3	1	Russian

The available eye fixation data sets in the literature

- **Our data set is the largest data globally and first eye movement data in Russian Language which covers appropriate age for identifying dyslexia**

Data

Dataset overview

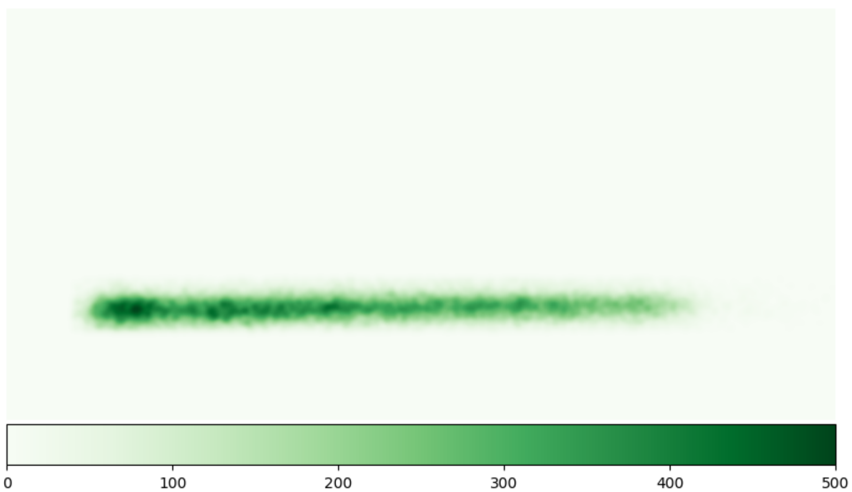
- The eye fixation data set consists of the following features:
  - **FIX_X and FIX_Y**: representing the fixation coordinates, alongs x-axis and y-axis, respectively.
  - **FIX_DURATION:** the duration (length) of each fixation time over a word in milliseconds.

Description	Quantity
Total Data Points	4298
Dyslexic Data Points	1462
Non-Dyslexic Data Points	2836

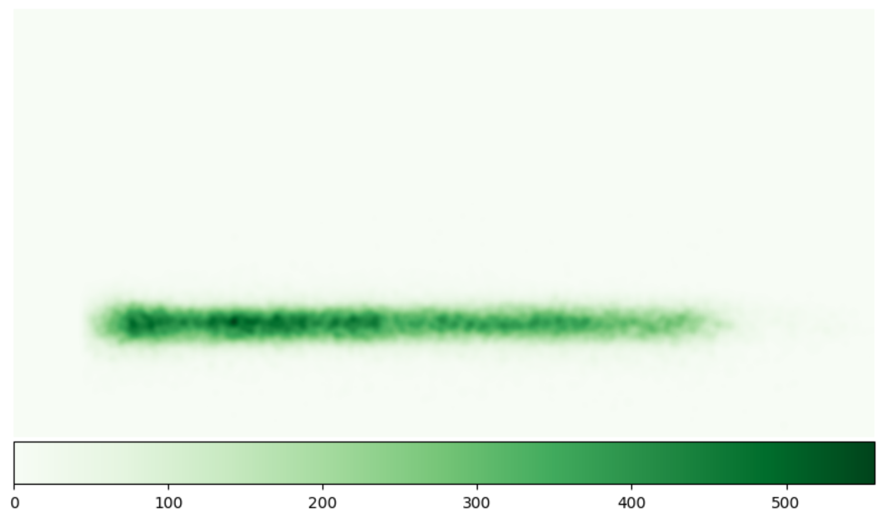
Group	Average Fixation Length (ms)
Overall	278.375 ± 218.225
Dyslexic Group	334.768 ± 224.311
Non-Dyslexic Group	247.531 ± 208.466

Data

Dataset overview



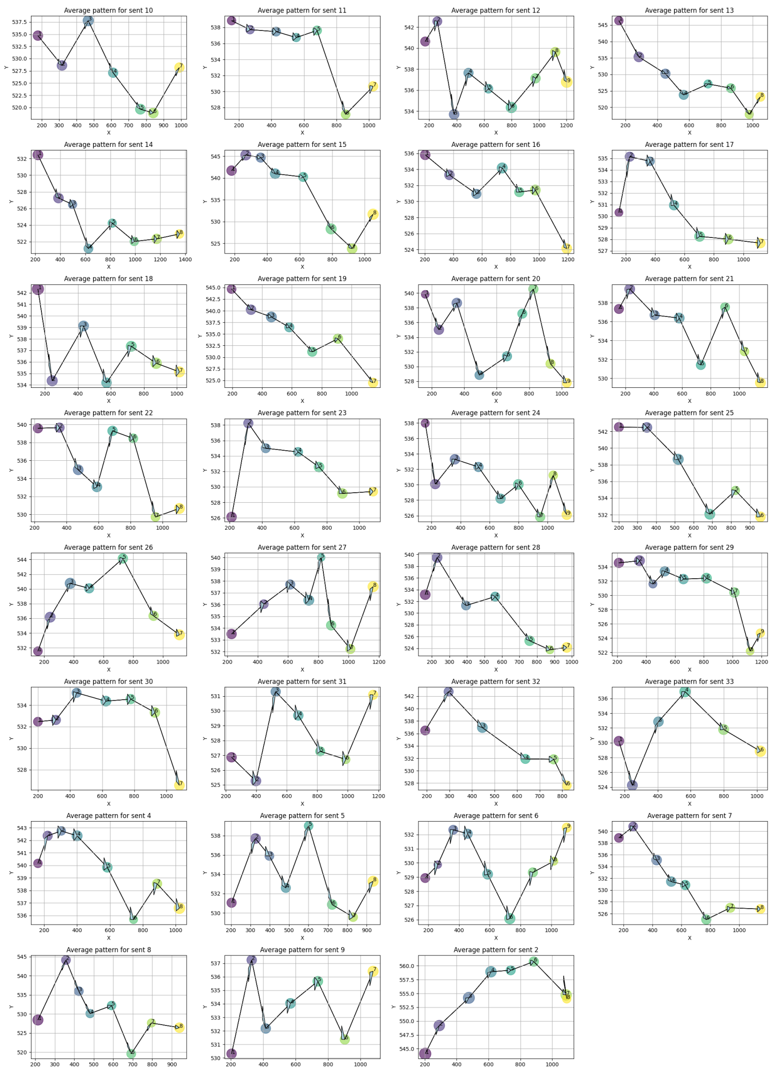
Dyslexic Heatmap



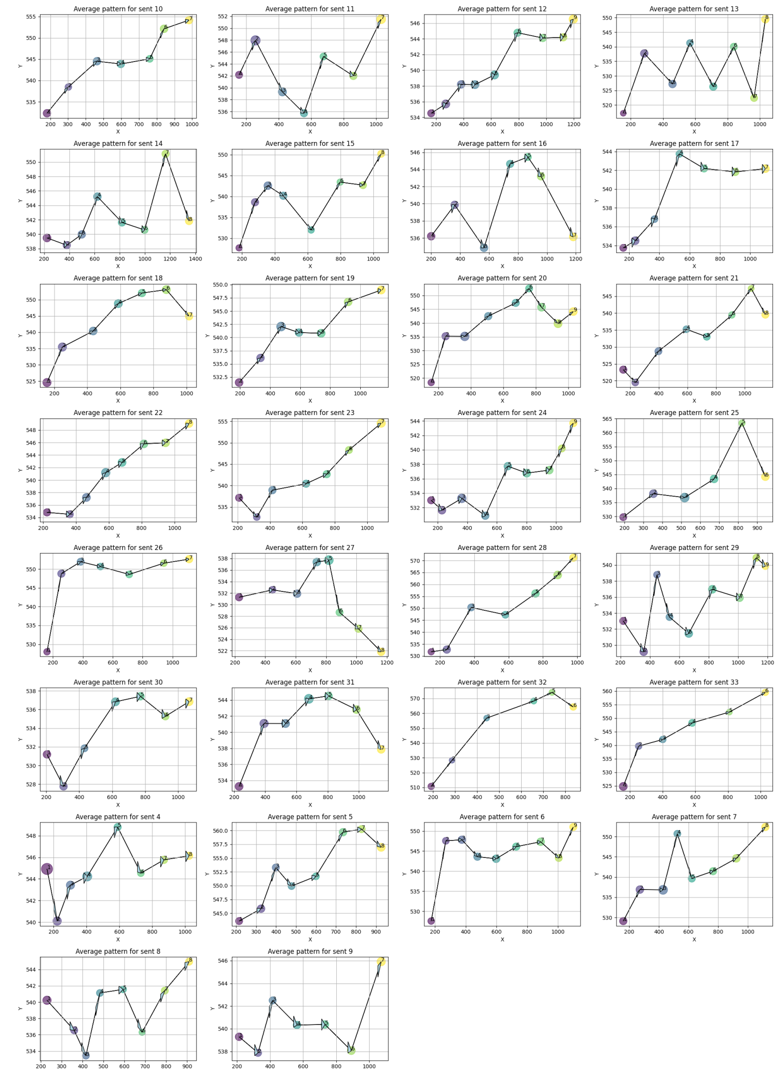
Non-Dyslexic Heatmap

- **Data point := (Fix_x, Fix_y) locate the marker and while the cumulative sum of fixation_duration forms their intensities**

NATIONAL RESEARCH UNIVERSITY



Dyslexic patterns of 3rd grade



Non-Dyslexic patterns of of 3rd grade

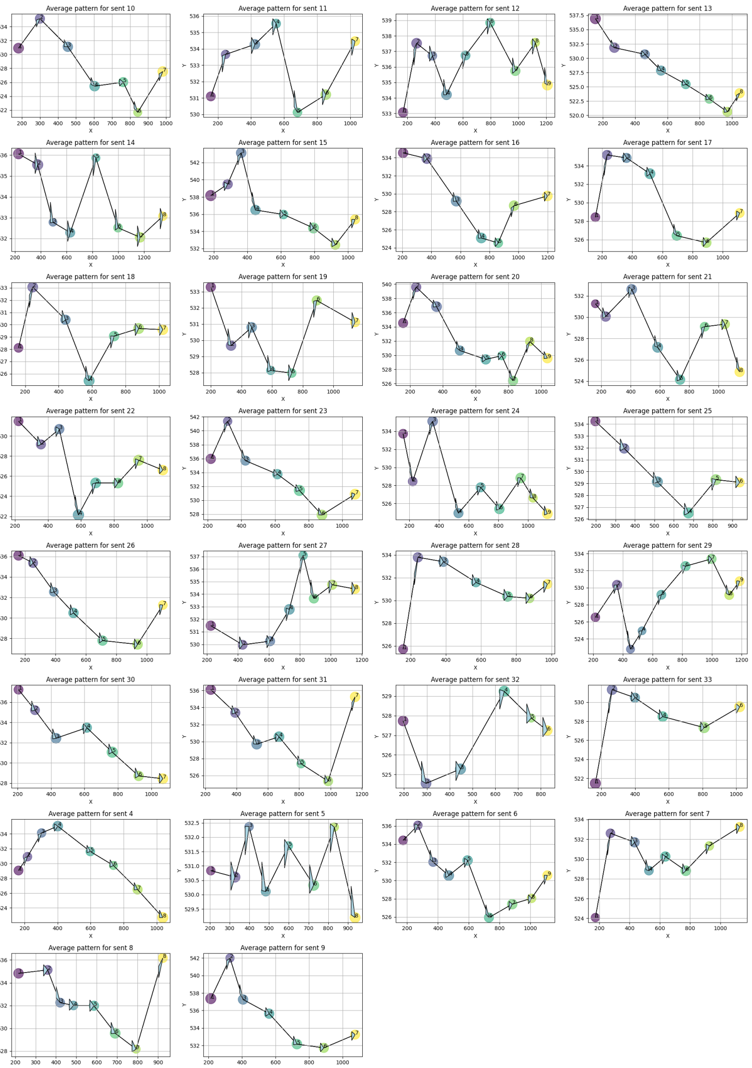
AICS

Laboratory of Artificial Intelligence for Cognitive Sciences

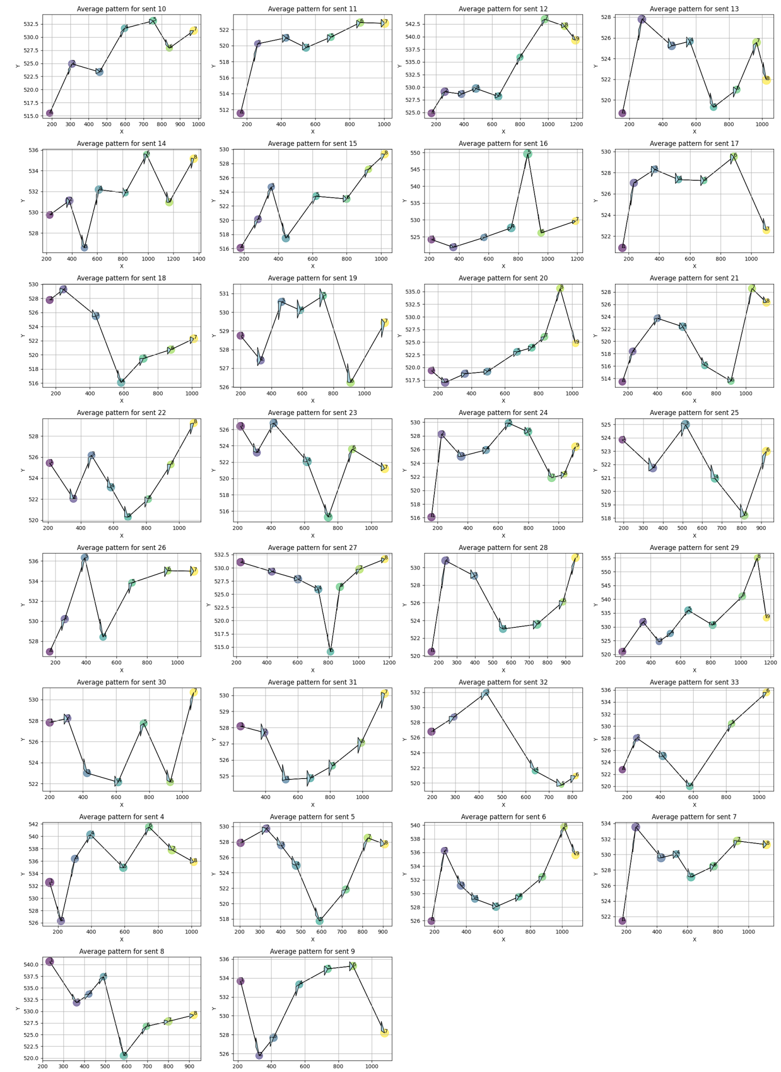
NATIONAL RESEARCH UNIVERSITY

AICS

Laboratory of Artificial Intelligence for Cognitive Sciences



Dyslexic patterns of 4th grade



Non-Dyslexic patterns of of 4th grade

Data

Data representations

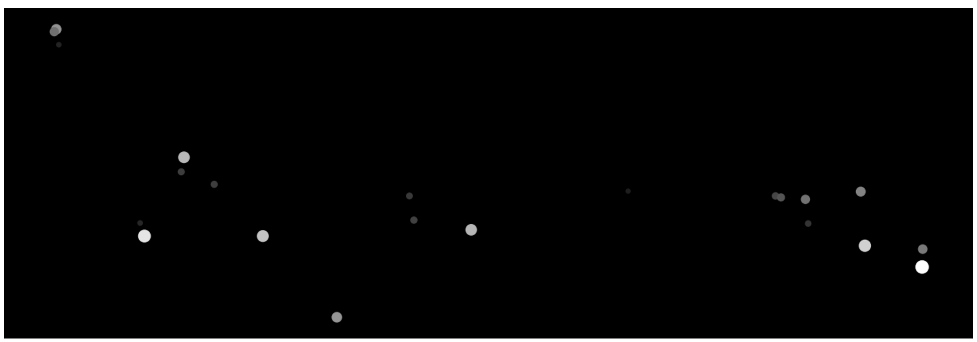
- **1D (windowed) data representation:** each window, containing 10 elements, the average word length of data set, consists of consecutive segments of the original series with a fixed number of consecutive elements.

Group	Average Number of Windows	Average Fix_x	Average Fix_y	Average Fix_dur
Overall	23	602	535	283
Dyslexic Group	24	597	532	333
Non-Dyslexic Group	22	605	537	257

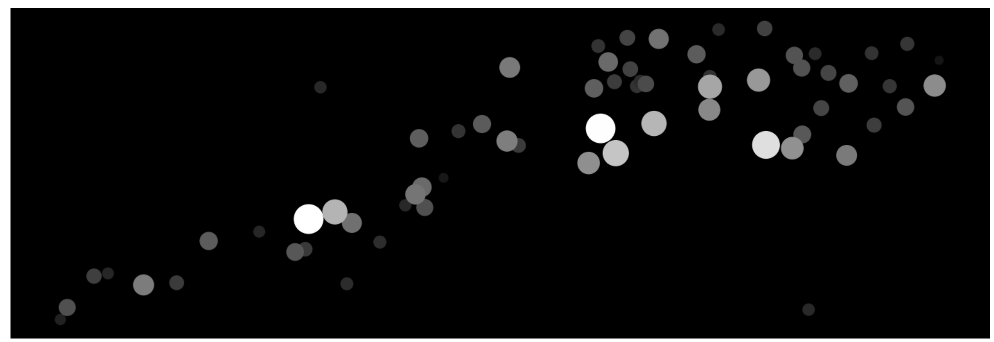
Data

Data representations

- **2D data representation:** eye-fixation plot.
  - Location of markers are determined by coordinates and intensity/size by fixation duration



Non-Dyslexic



Dyslexic

One can observe the significant difference between the eye movement patterns of participant with and without dyslexia: those with dyslexia fixation more and longer

Data

Data representations

- **3D data representation:**
    - Time-encoded marker-based:

$$\text{Marker size} \propto \text{Fixation Duration}$$

    - Trajectory tracking using connecting lines:

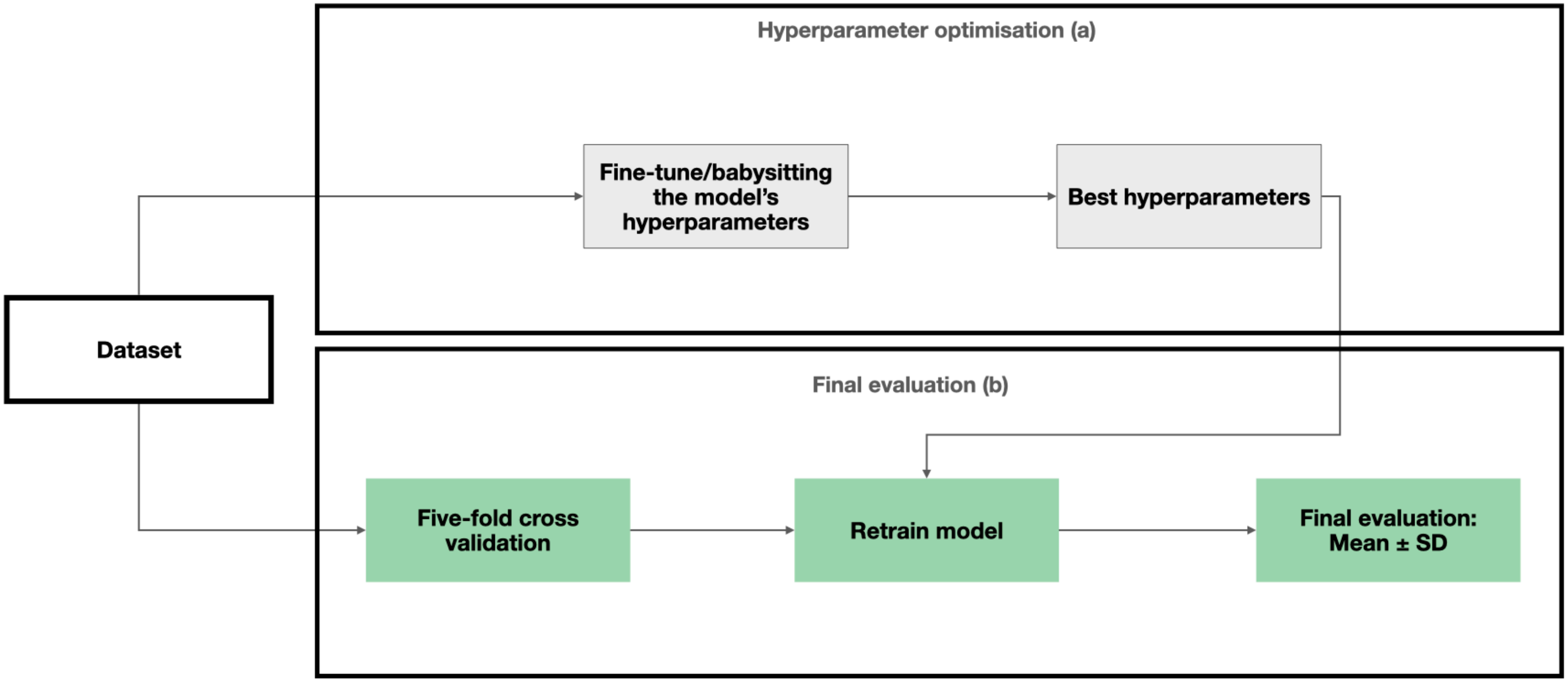
$$\text{Trajectory: Marker}_n \longrightarrow \text{Marker}_{n+1}$$

    - Multi-level markers:

$$\text{New level of markers} = \text{Previous level of markers} - \Delta z$$

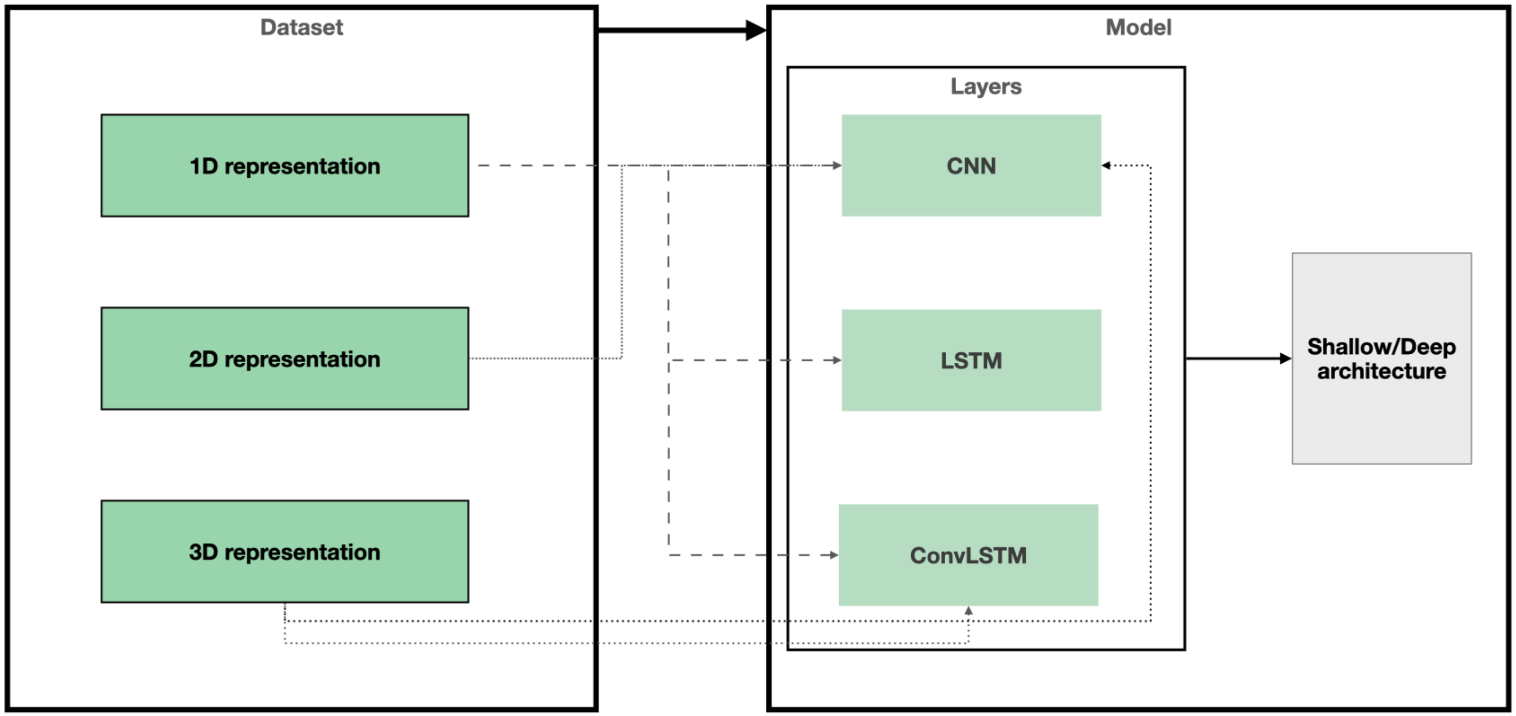
Methodology

Computational settings and hyperparameters tuning



Methodology

Computational settings and hyperparameters tuning



The deep architecture encompasses an expanded form of a shallow architecture, along with a custom-built ResNet model and a version of ResNet that has been pre-trained.

Methodology

Methods overview

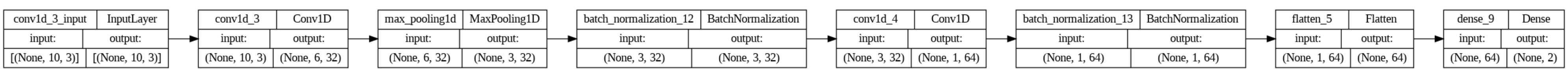
- **Deep Learning (DL) models:**
  - Long Short-Term Memory (LSTM): is a type of recurrent neural network (RNN) architecture used in deep learning for modeling sequential data, capable of learning long-term dependencies.
  - Convolutional Neural Networks (CNN): are a class of deep neural networks, primarily used in image recognition and processing, that are particularly adept at picking up patterns in spatial data through the use of convolutional layers.
  - Convolutional LSTM (ConvLSTM): is a type of recurrent neural network that combines convolutional layers with LSTM units, designed to capture both spatial and temporal patterns in data.
  - ResNet: is a type of CNN's that uses shortcut connections to skip one or more layers, significantly improving the ability to train deep networks by alleviating the vanishing gradient problem.
- **Ensemble learning (EL) models:**
  - Random Forest (RF): is an ensemble learning method that builds multiple decision trees for classification or regression tasks, and outputs the most common class or the average prediction.
  - Gradient Boosting Classifier (GBC): is a machine learning method that incrementally improves its predictions by correcting its own mistakes in a step-by-step manner, enhancing the accuracy of the model as it progresses.

Methodology

Methods

	Main layer type	Filters	Kernel size	LSTN units	Dense units	Dropout probobility
Shallow	Conv	[32; 256]	[3; 7]	–	–	–
	LSTM	–	–	[32; 256]	–	–
	ConvLSTM	[32; 256]	[3; 7]	–	–	[0.1; 0.5]
Deep	Conv	[16; 256]	[2; 7]	–	[64; 256]	[0.1; 0.5]
	LSTM	–	–	[16; 256]	[64; 256]	[0.1; 0.5]
	ConvLSTM	[16; 256]	[2; 7]	–	[64; 256]	[0.1; 0.5]

The search domain of the hyperparameters



Methodology

Evaluation metrics

- **Accuracy:** $$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$
- **Precision:** $$\text{Precision} = \frac{TP}{TP + FP}$$
- **Recall:** $$\text{Recall} = \frac{TP}{TP + FN}$$
- **F1-score:** $$F_1 = \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$
- Where TP, FP, TN and FN are the true positive, false positive, true negative and false negative respectively
- **ROC AUC:** The score is calculated from the area under the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings.
- All these metrics are bounded between zero and one, the closer to one the better.

Experimental results

1D results

	Type	Precision	Recall	F1 score	ROC-AUC score	Accuracy
Random prediction		0.590 ± 0.030	0.322 ± 0.017	0.385 ± 0.017	0.489 ± 0.031	0.322 ± 0.017
Shallow	Conv	0.658 ± 0.028	0.604 ± 0.069	0.626 ± 0.026	0.824 ± 0.003	0.743 ± 0.008
	LSTM	$\mathbf{0.774 \pm 0.026}$	$\mathbf{0.779 \pm 0.021}$	$\mathbf{0.777 \pm 0.023}$	$\mathbf{0.918 \pm 0.014}$	$\mathbf{0.839 \pm 0.017}$
	ConvLSTM	0.564 ± 0.021	0.591 ± 0.178	0.559 ± 0.096	0.764 ± 0.018	0.685 ± 0.015
Deep	Conv	0.684 ± 0.021	0.599 ± 0.041	0.637 ± 0.014	0.841 ± 0.005	0.755 ± 0.005
	LSTM	0.727 ± 0.033	0.748 ± 0.033	0.737 ± 0.032	0.893 ± 0.023	0.808 ± 0.024
	ConvLSTM	0.223 ± 0.133	0.006 ± 0.006	0.011 ± 0.012	0.486 ± 0.053	0.637 ± 0.002

Dyslexia classification using the 1D representation of eye fixation data with babysitted models: the average and standard deviation of five different data splits are reported and the winning results are bold-faced.

Shallow LSTM obtained the best and relatively acceptable results
The degeneration of its counterpart performance may have occurred due to incorrect tuning the hyperparameters

Experimental results

1D results

	Type	Precision	Recall	F1 score	ROC-AUC score	Accuracy
Random prediction		0.590 ± 0.030	0.322 ± 0.017	0.385 ± 0.017	0.489 ± 0.031	0.322 ± 0.017
Shallow	Conv	0.697 ± 0.028	0.598 ± 0.076	0.639 ± 0.035	0.845 ± 0.003	0.760 ± 0.006
	LSTM	0.859 ± 0.012	0.892 ± 0.020	0.875 ± 0.007	0.967 ± 0.003	0.909 ± 0.004
	ConvLSTM	0.478 ± 0.048	0.223 ± 0.261	0.225 ± 0.250	0.698 ± 0.038	0.647 ± 0.010
Deep	Conv	0.697 ± 0.028	0.598 ± 0.076	0.639 ± 0.035	0.845 ± 0.003	0.760 ± 0.006
	LSTM	$\mathbf{0.873 \pm 0.038}$	$\mathbf{0.884 \pm 0.016}$	$\mathbf{0.878 \pm 0.024}$	$\mathbf{0.968 \pm 0.009}$	$\mathbf{0.912 \pm 0.019}$
	ConvLSTM	0.548 ± 0.027	0.644 ± 0.146	0.582 ± 0.065	0.747 ± 0.010	0.678 ± 0.019
Ensemble	GBC	0.665 ± 0.008	0.651 ± 0.005	0.658 ± 0.005	0.845 ± 0.002	0.757 ± 0.003
	RF	0.677 ± 0.007	0.591 ± 0.009	0.631 ± 0.007	0.836 ± 0.002	0.752 ± 0.004

Dyslexia classification using the 1D representation of eye fixation data with tuned models: the average and standard deviation of five different data splits are reported and the winning results are bold-faced.

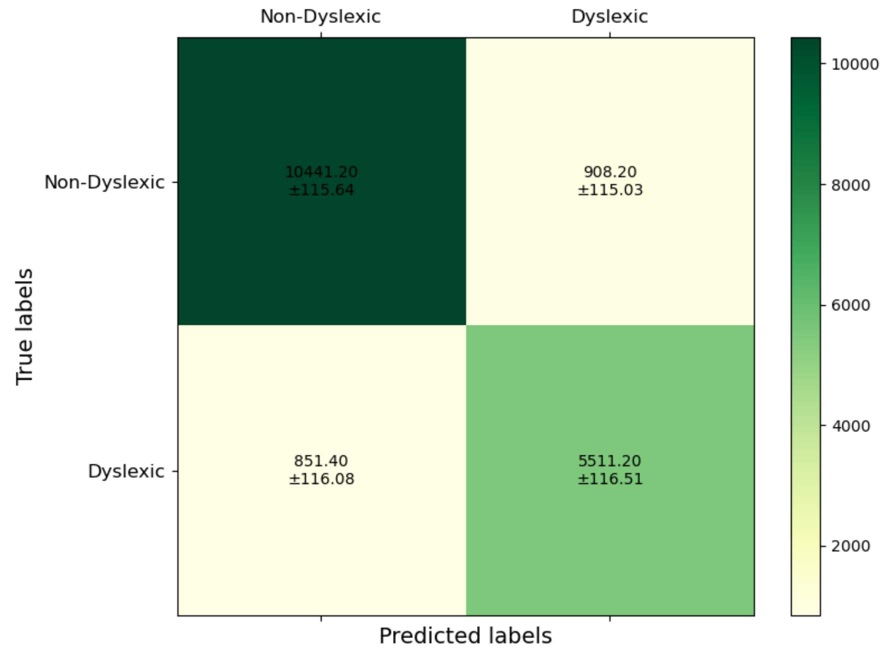
Deep LSTM obtained the best and completely acceptable results
For the ensemble models, a windowed version of the dataset was utilized. It underwent preprocessing, during which each window was flattened into a one-dimensional array.

Experimental results

1D results

Parameter	Value
$lstm_1$ units	64
$lstm_2$ units	64
$lstm_3$ units	96
$lstm_4$ units	160
$dense_1$ units	192
$dropout_1$	0.4
$dense_2$ units	128
$dropout_2$	0.2
epoch num.	150
lr	5.999×10^{-5}
optimizer	Adam

Hyperparameters of the tuned deep LSTM model for dyslexia classification using 1D representation.



Confusion matrix of the tuned deep LSTM model in classifying dyslexia using 1D representation: average and standart deviation for five different data splits.

In overwhelming majority of cases the model predicted the class labels correctly.

NATIONAL RESEARCH UNIVERSITY

Laboratory of Artificial Intelligence for Cognitive Sciences

Experimental results

2D results

	Type	Precision	Recall	F1 score	ROC-AUC score	Accuracy
Random prediction		0.590 ± 0.030	0.322 ± 0.017	0.385 ± 0.017	0.489 ± 0.031	0.322 ± 0.017
Babysitted	Conv shallow	0.540 ± 0.024	0.434 ± 0.035	0.480 ± 0.028	0.705 ± 0.022	0.682 ± 0.013
	Conv deep	0.595 ± 0.029	0.404 ± 0.0404	0.480 ± 0.034	0.723 ± 0.031	0.704 ± 0.014
	ResNet	0.633 ± 0.205	0.475 ± 0.391	0.382 ± 0.217	0.758 ± 0.018	0.639 ± 0.057
	ResNet pretrained	**0.659 ± 0.085**	**0.415 ± 0.151**	**0.482 ± 0.083**	**0.774 ± 0.012**	**0.715 ± 0.013**
Tuned	Conv shallow	0.592 ± 0.030	0.466 ± 0.047	0.521 ± 0.039	0.723 ± 0.019	0.710 ± 0.018
	Conv deep	0.604 ± 0.029	0.424 ± 0.045	0.496 ± 0.029	0.743 ± 0.011	0.709 ± 0.010

Dyslexia classification using the 2D representation of eye fixation data: the average and standard deviation of five different data splits are reported and the winning results are bold-faced.

ResNet is the winner of this setting

Experimental results

3D results

Architecture	Type	Visualization type	Precision	Recall	F1 score	ROC-AUC score	Accuracy
Random prediction			0.590 ± 0.030	0.322 ± 0.017	0.385 ± 0.017	0.489 ± 0.031	0.322 ± 0.017
Shallow	Conv	Time-encoded marker-based	0.419 ± 0.323	0.530 ± 0.437	0.310 ± 0.250	0.617 ± 0.009	0.525 ± 0.120
		Trajectory using lines	0.357 ± 0.199	0.310 ± 0.352	0.239 ± 0.230	0.612 ± 0.016	0.611 ± 0.067
		Multi-level cluster	0.381 ± 0.046	0.690 ± 0.355	0.435 ± 0.090	0.570 ± 0.030	0.469 ± 0.137
	ConvLSTM	Time-encoded marker-based	0.577 ± 0.220	0.354 ± 0.320	0.310 ± 0.181	0.629 ± 0.012	0.617 ± 0.076
		Trajectory using lines	**0.395 ± 0.035**	**0.794 ± 0.188**	**0.514 ± 0.021**	**0.643 ± 0.009**	**0.497 ± 0.091**
		Multi-level cluster	0.408 ± 0.038	0.593 ± 0.139	0.476 ± 0.044	0.613 ± 0.030	0.562 ± 0.064
Deep	Conv	Time-encoded marker-based	0.079 ± 0.158	0.131 ± 0.262	0.099 ± 0.197	0.595 ± 0.036	0.637 ± 0.047
		Trajectory using lines	0.242 ± 0.205	0.398 ± 0.475	0.220 ± 0.246	0.580 ± 0.033	0.551 ± 0.135
		Multi-level cluster	0.129 ± 0.159	0.309 ± 0.405	0.180 ± 0.224	0.513 ± 0.026	0.550 ± 0.138
	ConvLSTM	Time-encoded marker-based	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.5 ± 0.0	0.660 ± 0.001
		Trajectory using lines	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.486 ± 0.028	0.660 ± 0.001
		Multi-level cluster	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.5 ± 0.0	0.660 ± 0.001
ResNet	Conv	Time-encoded marker-based	0.211 ± 0.173	0.588 ± 0.481	0.311 ± 0.254	0.568 ± 0.110	0.490 ± 0.145
		Trajectory using lines	0.220 ± 0.182	0.566 ± 0.466	0.315 ± 0.258	0.610 ± 0.055	0.511 ± 0.144
		Multi-level cluster	0.068 ± 0.136	0.200 ± 0.400	0.102 ± 0.203	0.572 ± 0.046	0.596 ± 0.128
	ConvLSTM	Time-encoded marker-based	0.069 ± 0.137	0.2 ± 0.400	0.102 ± 0.204	0.591 ± 0.036	0.598 ± 0.125
		Trajectory using lines	0.273 ± 0.136	0.800 ± 0.400	0.407 ± 0.203	0.594 ± 0.034	0.406 ± 0.127
		Multi-level cluster	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.534 ± 0.034	0.660 ± 0.001

Dyslexia classification using the 3D representation of eye fixation data with babysitted models: the average and standard deviation of five different data splits are reported and the winning results are bold-faced.

Experimental results

3D results

Architecture	Type	Visualization type	Precision	Recall	F1 score	ROC-AUC score	Accuracy
Random prediction			0.590 ± 0.030	0.322 ± 0.017	0.385 ± 0.017	0.489 ± 0.031	0.322 ± 0.017
Shallow	Conv	Time-encoded marker-based	0.501 ± 0.106	0.359 ± 0.344	0.296 ± 0.201	0.636 ± 0.010	0.614 ± 0.059
		Trajectory using lines	0.444 ± 0.023	0.473 ± 0.107	0.451 ± 0.042	0.629 ± 0.013	0.616 ± 0.025
		Multi-level cluster	0.382 ± 0.036	0.741 ± 0.269	0.474 ± 0.075	0.618 ± 0.015	0.477 ± 0.112
	ConvLSTM	Time-encoded marker-based	0.441 ± 0.061	0.496 ± 0.372	0.372 ± 0.160	0.618 ± 0.013	0.569 ± 0.107
		Trajectory using lines	0.548 ± 0.252	0.564 ± 0.443	0.335 ± 0.225	0.630 ± 0.014	0.518 ± 0.128
		Multi-level cluster	0.382 ± 0.036	0.741 ± 0.269	0.474 ± 0.075	0.618 ± 0.015	0.477 ± 0.112
Deep	Conv	Time-encoded marker-based	**0.517 ± 0.052**	**0.410 ± 0.183**	**0.421 ± 0.124**	**0.675 ± 0.007**	**0.656 ± 0.020**
		Trajectory using lines	0.275 ± 0.226	0.045 ± 0.060	0.072 ± 0.091	0.637 ± 0.016	0.656 ± 0.004
		Multi-level cluster	0.15 ± 0.3	0.005 ± 0.009	0.009 ± 0.018	0.632 ± 0.015	0.661 ± 0.002
	ConvLSTM	Time-encoded marker-based	0.284 ± 0.233	0.028 ± 0.049	0.046 ± 0.077	0.609 ± 0.017	0.659 ± 0.001
		Trajectory using lines	0.119 ± 0.238	0.030 ± 0.060	0.048 ± 0.095	0.593 ± 0.059	0.663 ± 0.007
		Multi-level cluster	0.086 ± 0.171	0.005 ± 0.010	0.009 ± 0.018	0.524 ± 0.024	0.659 ± 0.001

Dyslexia classification using the 3D representation of eye fixation data with tuned models: the average and standard deviation of five different data splits are reported and the winning results are bold-faced.

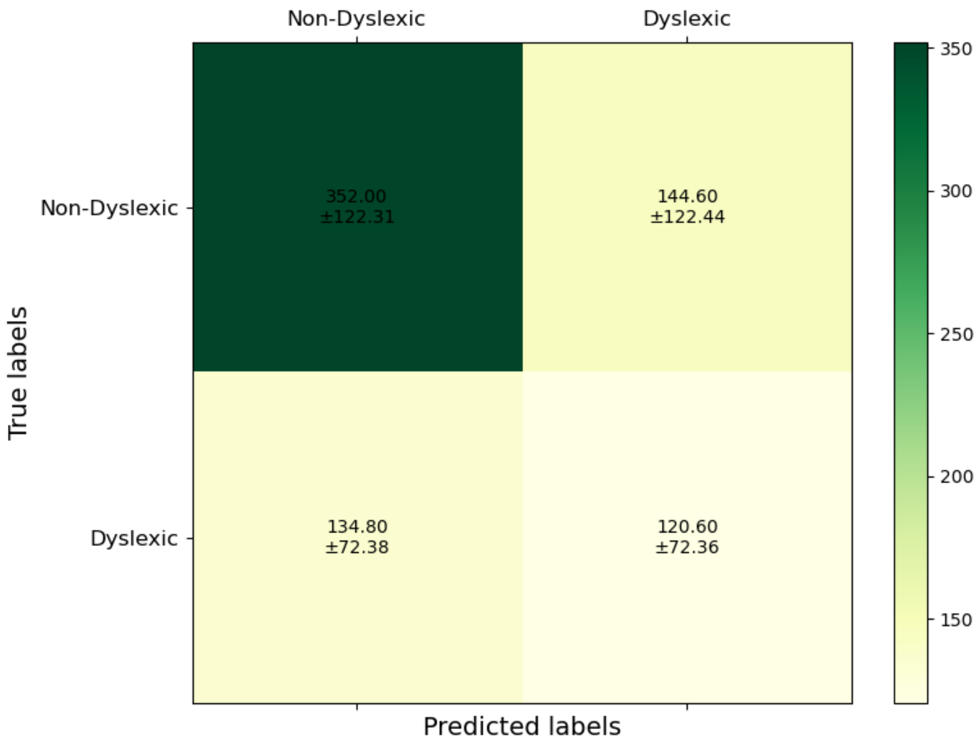
None of the models obtained completely acceptable results, which may imply the lack of
(i) existence of any meaningful relations in the between-word eye fixations; (ii) data points to train models for such a complex data structure

Experimental results

3D results

Parameter	Value
conv_1 filters	64
conv_1 kernels	7
conv_2 filters	96
conv_2 kernels	7
conv_3 filters	128
conv_3 kernels	3
conv_4 filters	96
conv_4 kernels	3
dense_1 units	256
dropout_1	0.1
dense_2 units	32
dropout_2	0.2
epoch num.	35
lr	1×10^{-5}
optimizer	Adam

Hyperparameters of the tuned deep Conv model using 3D representation



Confusion matrix of the tuned deep Conv model in classifying dyslexia using 3D representation with time-encoded marker-based version: average and standart deviation for five different data splits.

Conclusion and future work

- **Conclusion:**
  - We scrutinized three way of representation eye fixation data for training AI models.
  - All questions were successfully answered:
    - Q1: using only eye fixation data, we achieved reliable and accurate results in predicting dyslexia.
    - Q2: eye-fixation data, rich in details like eye position and fixation duration, reveals distinct movement patterns that differentiate children with dyslexia from those without, making it valuable for artificial intelligence models.
    - Q3: utilizing one-dimensional windowed representations, two-dimensional fixation graphs, and enhanced three-dimensional fixation graphs with a temporal dimension.
    - Q4: the best models for analyzing eye fixation data are those based on LSTM layers, with Deep LSTM outperforming all others in every evaluated parameter.

Conclusion and future work

- **Future work:**
  - Further investigation to study the performance of models for 3D data representation.
  - Using more complex deep learning algorithms, such as vision transformers.
  - Running clinical trials.
  - Research interpreting the performance of our best model on 1D data representation.

Thank you!

Soroosh Shalileh

@SSHALILEH

sr.shalileh@gmail.com

Maria Krylova

@M_VLKR

m.vlkrylova@gmail.com