Towards a robust and accurate screening tool for dyslexia with data augmentation using GANs

Thomais Asvestopoulou [1,2,^], **Victoria Manousaki** [1,2], **Antonis Psistakis** [1,2], **Erjona Nikolli** [1,2], **Vassilios Andreadakis** [3], **Ioannis M. Aslanides**[4], **Yannis Pantazis**[5], **Ioannis Smyrnakis** [2,3] and **Maria Papadopouli** [1,2]

[1]Department of Computer Science, University of Crete, Heraklion, Greece

[2] Institute of Computer Science, Foundation for Research and Technology-Hellas, Heraklion, Greece

[3] Optotech Ltd., Heraklion, Greece

[4] Emmetropia Eye Institute, Heraklion, Greece

[5] Institute of Applied and Computational Mathematics, Foundation for Research and Technology-Hellas, Heraklion, Greece

^ First author: Thomais Asvestopoulou, M.Sc. student, email: tasvest@ics.forth.gr

Presenting author: Yannis Pantazis , email: pantazis@iacm.forth.gr

* Corresponding author: Maria Papadopouli, email: mgp@ics.forth.gr

ABSTRACT

Eye movements during text reading can provide insights about reading disorders. We developed the DysLexML, a screening tool for developmental dyslexia, based on various machine learning algorithms that analyze gaze points recorded via eye-tracking during silent reading of children. We comparatively evaluated its performance using measurements collected from two systematic field studies with 221 participants. This work presents DysLexML and evaluates its performance. It identifies the features with prominent predictive power and performs dimensionality reduction. Specifically, DysLexML achieves its best performance using linear SVM, with an accuracy of 97% and 84% respectively, using a small feature set. We show that DysLexML is also robust in the presence of noise in the form of small displacements of the fixation positions. These encouraging results set the basis for developing screening tools in less controlled, larger-scale environments, with inexpensive commercial eye-trackers, potentially reaching a larger population for early intervention and potentially larger social impact. Unlike other related studies, DysLexML achieves the aforementioned performance by employing only a small number of selected features that have been identified with prominent predictive power. Finally, we developed a new data augmentation/substitution technique based on GANs for generating synthetic data similar to the original distributions. The increase of the available data enables the use of more sophisticated ML and deep learning algorithms, which may further increase the performance.

REFERENCES

[1] T. Asvestopoulou, V. Manousaki, A. Psistakis, E. Nikolli, V. Andreadakis, I. M. Aslanides, Y. Pantazis, I. Smyrnakis and M. Papadopouli. 2019. “Towards a robust and accurate screening tool for dyslexia with data augmentation using GANs.”, *19th IEEE International Conference on BioInformatics and BioEngineering (BIBE)*

ACKNOWLEDGEMENTS

This work has been partially funded from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT) under grant agreement No. 2285, the Erasmus+ International Mobility between University of Crete and Harvard Medical School 2017-1-EL01-KA107-035639, the Marie Curie RISE NHQWAVE project under grant agreement No. 4500, and the Human Resources Development, Education and Life Lifelong Learning for the implementation of the European Social Fund and the Youth Employment Initiative.