Towards Detecting Dyslexia in Children's Handwriting Using Neural Networks

Katie Spoon [1]   **David Crandall** [1]   **Katie Siek** [1]

Abstract

Literacy is the most reliable indicator for future success (Ritchie & Bates, 2013). Dyslexia is a learning disability that hinders a person's ability to read (Learning Disabilities Association of America). Dyslexia needs to be caught early, however, teachers are not trained to detect dyslexia (Walsh et al., 2006) and screening tests are used inconsistently. We propose (1) two new data sets of handwriting collected from children with and without dyslexia, and (2) an automated early screening technique to be used in conjunction with current approaches, to accelerate the detection process.

1. Introduction

An estimated 20% of the U.S. population has dyslexia or a similar language-based learning disability (National Center for Learning Disabilities, 2017a)—that's 11.32 million kids (National Center for Education Statistics, 2018). Dyslexia is not tied to IQ (Gabrieli et al., 2011); many smart, successful people have dyslexia. Students with dyslexia can perform well if they receive the interventions and accommodations they need. If a child struggles to read in third grade, they are four times more likely to drop out of high school (Hernandez, 2011), so the recommended age for detection is third grade. However, fewer than 5% of students with dyslexia are actually detected and/or diagnosed by third grade (National Center for Learning Disabilities, 2017b) The detection process is often not straightforward, but an obstacle course of hoops and roadblocks, often delaying diagnosis and causing critical years of learning to be lost.

Moreover, this problem has not entered the public consciousness: according to a recent survey (Cortiella & Horowitz, 2014), 53% of people think that learning disabilities are diagnosed in grades 1-4, and an additional 23% think they are diagnosed in kindergarten. But most detection comes much later, to the detriment of the student and, by extension, society. Teachers are often blamed for this problem, but teachers do not routinely have the training to detect learning disabilities (Walsh et al., 2006). There is thus a strong need for earlier, easier, and less costly detection of dyslexia.

In this paper, we present ongoing work in which we explore the potential of modern machine learning methods to automatically identify possible indications of dyslexia in handwriting. In particular, we propose a multi-stream convolutional neural network to classify handwriting as suggestive of dyslexia or not, with the intention of creating a triage tool that could, together with evidence from teachers and parents, be used to refer the child to a school psychologist for further testing.

2. Motivation

2.1. Related Work

As dyslexia and other learning disabilities have gained awareness, various approaches have been used to try to detect and diagnose these disabilities.

Conventional detection of dyslexia often focuses on behavioral aspects, including reading and writing proficiency, IQ, phonological awareness (sounding out words), and working memory, typically all assessed using standardized tests. This process is time-consuming, and can miss many cases of dyslexia since every person experiences it differently. Many children with dyslexia perform well on standardized tests by overcompensating, as many are at average or above average intelligence.

Many groups have used machine learning to detect dyslexia using students' test scores as well as demographic or survey data as features (Kohli & Prasad, 2010). Loizou and Laouris used the results of four tests for dyslexia as features in several machine learning techniques to diagnose students, amounting to 226 total features (2011). Costa et al. used a total of 144 features, including some hand-collected from interviews with families and students (asking about discipline, liking school, having friends, etc.), to assess learning disabil-

---

[1]School of Informatics, Computing and Engineering, Indiana University, Bloomington, IN, USA. Correspondence to: Katie Spoon <kspoon@iu.edu>.

ities (2013). Others have successfully diagnosed dyslexia from computational classification of brain electrical activity (Duffy et al., 1980; Lehongre et al., 2013) or by tracking eye movements (De Luca et al., 2002). Very recent work by Rezvani et al. used machine learning to automate EEG (brain wave) and brain image analysis to potentially decrease the cost of brain screening tests (2019). Other work even combines EEG tests and genetic tests to diagnose dyslexia (Wilcke et al., 2015). Gathering hundreds of these high-level features or administering neurological tests is expensive and time-consuming. Here we consider the complimentary goal of detecting potential cases of dyslexia automatically, leaving it to experts or more extensive automated tests to confirm a diagnosis.

A new wave of interactive technology to detect dyslexia has emerged in recent years. Rello et al. designed a game that can detect dyslexia across languages (2016), only requiring a student to play the game for approximately 15 minutes to calculate a likelihood of the student having dyslexia. Other research groups have also designed games for the detection of dyslexia (Ekhsan et al., 2012; Bartolome et al., 2012; Gaggi et al., 2012).

However, none of these applications has attempted to detect dyslexia through handwriting. Handwriting is easy to collect but still could provide substantial evidence for making a dyslexia detection decision. Decades of computer vision research have allowed Optical Character Recognition (OCR) to reach nearly perfect performance, but the vast majority of this work is for adult handwriting. Janet Read has been studying handwriting recognition for children for many years. Most of her work has been centered on designing systems and environments for children to improve their handwriting (Read & Horton, 2004), so recognition of children's handwriting was an important component. Due to the lack of a large, publicly available data set of children's handwriting, she used human-in-the-loop techniques, and recommended the collection of a larger data set. Other groups have also investigated handwriting recognition to automate scoring standardized tests (Srihari et al., 2008).

In this paper, we propose using children's unstructured handwriting samples as a technique for early detection of dyslexia. To our knowledge, we are the first to propose this kind of detection system.

3. Data Collection

We contribute a new data set of children's handwriting from grades K-6, partitioned in two:

1. An unconstrained, "messy" set of creative handwriting, where parents were asked to upload a photo of handwriting their child had already written. We currently have 2 samples from students with dyslexia and 15
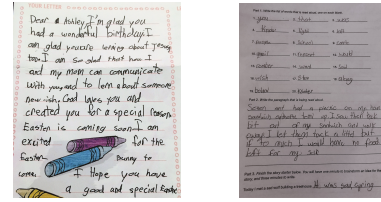

Figure 1. (a) the messy data set contains handwriting that was written for school or for fun, and is not standardized, but in (b) the controlled data set contains handwriting that is the same for everyone

samples from students without dyslexia.

2. A controlled data set, collected by children who were asked to write down words and a paragraph read by parents or researchers. We currently have 9 samples from students with dyslexia and 62 samples from students without dyslexia.

Each sample of the controlled data set contains three main components:

1. Twenty words that are typically used to screen for dyslexia when tests are administered. These are common sight words that are easy for children without dyslexia to learn but difficult for children with dyslexia to learn, and especially to spell. (Kindergarten and first grade were only asked to write 10 words.)

2. A paragraph that the children were asked to write. This is used to determine how well children can transition between words, and if it is easier for them to spell words if they are in sentence form. (Kindergarten and first grade only wrote one sentence.)

3. Children were given a story starter (*Today I met a sad wolf building a tree house...*) and asked to complete the story in three minutes. Kindergarten and first grade students did not complete this part.

Data collection will continue until we have at least 50 samples from students with dyslexia and 100 samples from students without dyslexia for both data sets. Samples from students without dyslexia are much easier to obtain because students with diagnosed dyslexia before sixth grade are rare — something our project is trying to help solve!

We collected the controlled data set (the second partition) through two main IRB-approved approaches:

1. **Classroom collection**: we taught a lesson plan to several classrooms of students to collect standardized data.

2. **Parental collection**: through an online form, parents were invited to upload handwriting their children had already written for the messy data set. They were also invited to print the form, administer the longer study, and upload that data to the controlled data set.

We collected demographic data along with the handwriting, including grade, age, gender, hand the child wrote with, whether or not they had dyslexia (and if so, when it was detected as well as diagnosed), ethnicity, whether they are hispanic or latino, highest education of their parents, income, marital status, other languages they speak, how often they were read to ages 0-5, how much they enjoy school, and how much they enjoy writing.

4. Our approach

We initially tried to use Optical Character Recognition (OCR) software (e.g., Tesseract (Google)) to transcribe the writing in the images, and then use machine learning to predict dyslexia based on spelling errors and other features of the text. This approach did not work well due to the variability in the handwriting, so instead we used visual features of the writing itself.

We based our technique on an existing approach for handwriting recognition (Dwivedi, 2018; Xing & Qiao, 2016) that tried to match different writers to their handwriting. We modified it to assume that there were only two different writers: one with dyslexia, and one without. We ran this approach on the controlled data set.

4.1. Preprocessing

We split each image into lines of text and then generated 50 random patches per line. To split the images into lines of text, we used Arvanitopoulos & Süsstrunk's seam carving technique for historical manuscripts (2014), which – like children's handwriting – can be quite messy. Even with their approach, some of the samples were still too challenging and needed to be fixed manually (which, in practice, could be easily done with minimal human effort).

After this extraction process, we had 1060 total words from the first section of the data form. Additionally, from the second section of the form we had 222 lines of text, and from the third section of the form we had 185 lines of text. Once we had extracted these text lines, we cropped them, resized to a height of 113 pixels, and generated 50 random 113x113 patches for each line. For the words, we generated only 5 random patches per word, amounting to 5,300 patches of singular words (section 1 of the document in Figure 1b). We then cropped the lines of text, which gave us 11,100 patches from the second section and 9,250 patches from the third section. Using randomized patches instead of individual

letters is attractive because it is language-agnostic, and we plan to expand this project to other languages in future work.

4.2. Network Architecture

We applied a convolutional neural network (CNN) to this task, and implemented it using Keras and TensorFlow. The network has 5 convolutional layers, 3 max-pooling (MP) layers, 2 fully-connected (FC) layers, and a dropout layer. We experimented with a variety of batch sizes (1, 4, 16, 32), as well as number of patches per line (5, 10, 25, 50). We split the data into train:validation:test in a 4:1:1 split. The highest accuracy we obtained can be seen in Section 5, with a batch size of 4 and the number of patches per line set to 50 (most likely due to the small amount of data we have).

5. Preliminary Results

We ran the experiment on all grades, but specifically chose to focus on the results from students in third grade. Using five-fold cross validation, we obtained an average accuracy of $55.7\pm1.4\%$ compared to a random baseline of 50%.

More data is required to study the results further, especially more data from students with dyslexia. However, these preliminary results show promise, as they are much higher than the current detection rate of teachers and parents by third grade.

6. Implications & Ongoing Work

This data set is still growing by the day, so our immediate goal is to collect a larger data set. With the few samples we have, overfitting is a concern so it is difficult to accurately interpret and utilize the results. We have recently added gift cards as incentives for our study and we have a growing interest from parents in participation.

In future work, we plan on trying unsupervised approaches including clustering to group the data into clusters that could then be inspected by human experts. Due to the prevalence of dyslexia and the number of cases that go undiagosed, particularly in elementary school, it is possible that some of the negative examples in our training data set may actually be from children with dyslexia. Clustering or other unsupervised approaches may provide us with more information about what handwriting from the 20% looks like.

Additionally, we would like to visualize the features the neural network is utilizing. Unfortunately, neural networks act as a *black box*, taking in data and producing an output, without much explanation. In order for a system like this to be considered, educators, administrators and parents need to have an idea of the types of features the network used to determine whether or not the child could potentially have dyslexia. We intentionally collected a large amount of de-

mographic data with the new data sets in order to graph the results and make sure our model is not accidentally leaving out an entire gender, race, socioeconomic level, etc. This will be extremely important as we collect more data and as the accuracy of the model increases.

The implications of this system reach farther than aiding teachers; many parents have struggled to advocate for their children within the public school system. Indeed, Sandman-Hurley, author of an advocacy book for parents of children with dyslexia within the public school system, *Dyslexia Advocate!*, sternly warns parents: "You need to know what to expect, what education rights are afforded you by the 2004 IDEA [Individuals with Disabilities Education Act], and how to prepare for any obstacles. Prepare to become **'that parent'**." In addition, once children start the diagnosis process, parents are asked to provide samples of their children's handwriting from as many years as the child has been writing. Thus, this system would have the potential to show a child's handwriting over time.

Parents should not have to be able to read and understand complicated laws in order for their child to have a free appropriate public education. An automated piece in the dyslexia detection process could potentially make sure every child is afforded the same opportunities for success, no matter their background.

References

Arvanitopoulos, N. and Ssstrunk, S. Seam Carving for Text Line Extraction on Color and Grayscale Historical Manuscripts. volume 14 of *International Conference on Frontiers in Handwriting Recognition*, pp. 726–731, Heraklion, Greece, 2014.

Bartolome, N., Zorrilla, A., and Zapirain, B. Dyslexia Diagnosis in Reading Stage through the Use of Games at School. volume 17 of *International Conference on Computer Games*, pp. 12–17, 2012.

Cortiella, C. and Horowitz, S. H. The State of Learning Disabilities: Facts, Trends and Emerging Issues. New York: National Center for Learning Disabilities, 2014.

Costa, M., Zavaleta, J., Manuel Serra da Cruz, S., Manhes, M., Cerceau, R., Alfredo Carvalho, L., and Mousinho, R. A Computational Approach for Screening Dyslexia. volume 26 of *IEEE International Symposium on Computer-Based Medical Systems (CBMs)*, Porto, Portugal, 2013.

De Luca, M., Borrelli, M., Judica, A., Spinelli, D., and Zoccolotti, P. Reading Words and Psuedowords: An Eye Movement Study of Developmental Dyslexia, 2002.

Duffy, F., Denckla, M., Bartels, P., Sandini, G., and Kiessling, L. Dyslexia: Automated Diagnosis by Comput-erized Classification of Brain Electrical Activity. *Annals of Neurology*, 7:421–428, 1980.

Dwivedi, P. Handwriting Recognition Using TensorFlow and Keras. https://towardsdatascience.com/handwriting-recognition-using-tensorflow-and-kera 2018.

Ekhsan, H., Ahmad, S., Halim, S., Hamid, J., and Mansor, N. The Implementation of Interactive Multimedia in Early Screening of Dyslexia. International Conference on Innovation Management and Technology Research (ICIMTR), pp. 566–569, 2012.

Gabrieli, J., Hoeft, F., Tanaka, H., Black, J. M., Stanley, L. M., Kesler, S. R., Reiss, A. L., Hulme, C., and Whitfield-Gabrieli, S. The Brain Basis of the Phonological Deficit in Dyslexia is Independent of IQ. *Psychological Science*, 22:1442–1451, 2011.

Gaggi, O., Galiazzo, G., Palazzi, C., Facoetti, A., and Franceschini, S. A Serious Game for Predicting the Risk of Developmental Dyslexia in Pre-Readers Children. volume 21 of *International Conference on Computer Communications and Networks (ICCCN)*, pp. 1–5, 2012.

Google. Tesseract. Available at https://github.com/tesseract-ocr/.

Hernandez, D. Double Jeopardy: How Third-Grade Reading Skills and Poverty Influence High School Graduation. *Annie E. Casey Foundation*, 2011.

Kohli, M. and Prasad, T. Identifying Dyslexic Students by Using Artificial Neural Networks. volume 1 of *World Congress on Engineering (WCE)*, London, U.K., 2010.

Learning Disabilities Association of America. Dyslexia. https://ldaamerica.org/types-of-learningdisabilities/dyslexia/.

Lehongre, K., Morillon, B., Giraud, A.-L., and Ramus, F. Impaired Auditory Sampling in Dyslexia: Further Evidence from Combined fMRI and EEG. *Frontiers in Human Neuroscience*, 7:454, 2013.

Loizou, A. and Laouris, Y. Developing Prognosis Tools to Identify Learning Difficulties in Children Using Machine Learning Technologies. *Cognitive Computation*, 3:490–500, 2011.

National Center for Education Statistics. Back to School Statistics, 2018.

National Center for Learning Disabilities. The State of LD: Understanding the 1 in 5.

https://www.ncld.org/archives/blog/
the-state-of-ld-understanding-the-1-in-5,
2017a.

National Center for Learning Disabilities. Identifying
Struggling Students. https://www.ncld.org/
identifying-struggling-students, 2017b.

Read, J. and Horton, M. Demonstrating CobWeb - An
Innovative Writing Environment for Children. World
Conference on Educational Multimedia, Hypermedia &
Telecommunications, pp. 1908–1910, Lugano, Switzer-
land, 2004.

Rello, L., Williams, K., Ali, A., White, N., and Bingham,
J. Dytective: Towards Detecting Dyslexia Across Lan-
guages Using an Online Game. volume 13 of *Web for All
Conference*, Montreal, CA, 2016.

Rezvani, Z., Zare, M., Zaric, G., Bonte, M., Tijms, J.,
Van der Molen, M., and Fraga Gonzolez, G. Ma-
chine Learning Classification of Dyslexic Children based
on EEG Local Network Features. *BioRxiv 569996
[Preprint]*, 2019.

Ritchie, S. and Bates, T. Enduring Links from Child-
hood Mathematics and Reading Achievement to Adult
Socioeconomic Status. *Psychological Science*, 24(7):
1301–1308, 2013.

Srihari, S., Collins, J., Srihari, R., Srinivasan, H., Shetty,
S., and Brutt-Griffler, J. Automatic Scoring of Short
Handwritten Essays in Reading Comprehension Tests.
Artificial Intelligence, 172, 2008.

Walsh, K., Glaser, D., and Wilcox, D. What Education
Schools Arent Teaching About Reading and What Ele-
mentary Teachers Arent Learning. *National Council on
Teacher Quality*, 2006.

Wilcke, A., Mller, B., Schaadt, G., Kirsten, H., and Boltze, J.
High Acceptance of an Early Dyslexia Screening Test In-
volving Genetic Analyses in Germany. *European Journal
Of Human Genetics*, 24, 2015.

Xing, L. and Qiao, Y. DeepWriter: A Multi-stream Deep
CNN for Text-Independent Writer Identification. vol-
ume 15 of *International Conference on Frontiers in Hand-
writing Recognition (ICFHR)*, pp. 584–589, Shenzhen,
China, 2016.