Can We (and *Should We*) Use AI to Detect Dyslexia in Children's Handwriting?

Katie Spoon
Indiana University Bloomington
kspoon@iu.edu

David Crandall
Indiana University Bloomington
djcran@indiana.edu

Katie Siek
Indiana University Bloomington
ksiek@indiana.edu

Marlyssa Fillmore
The Academic Achievement Center
Columbus, IN

Abstract

Reading is a critical skill that affects the development of all other skills, as well as self-esteem. Language-based learning disabilities like *dyslexia* interfere with a child's ability to read. This paper follows up on the previous proof-of-concept research of Spoon et al. [1], who developed a system that used computer vision to classify handwriting samples as indicative of dyslexia or not. In this paper, we contribute (1) a subset of a robust data set from a pilot school system and (2) visual and quantitative evidence that suggests dyslexia detection, eliminating confounding factors. Here we seek to show that not only is it possible to classify handwriting as characteristic of dyslexia, we can now explain the behavior of the neural network to the end users: teachers, parents, and the kids affected by those decisions.

1 Introduction

Literacy is deeply connected to a child's outcomes in life [2]—in employment, health, safety and more, reading is crucial. This is part of why a *Quality Education* for every child in the world is one of the U.N. Sustainable Development Goals [3]. Dyslexia is a learning disability that hinders a person's ability to read [4]. It is critical that dyslexia is diagnosed early, however, teachers are not trained to detect dyslexia [5] and screening tests are used inconsistently.

While up to 20% of the U.S. population is estimated to have dyslexia or another language-based learning disability [6], very few of them are detected with the disability by the appropriate age. A child that struggles to read in third grade is four times more likely to drop out of high school [7], but fewer than 5% of students with dyslexia are actually detected by the critical age of third grade [8].

Most people (**76%**, according to a recent survey [9]) believe that learning disabilities are diagnosed before fourth grade. But most detection comes much later or not at all, delaying critical years of learning for millions of students [10] and costing our society billions [11]. Teachers do not have the time or the training to detect the subtleties of dyslexia along with their immense workload. There is a strong need for earlier, easier, and less costly detection of dyslexia.

In this paper, we build upon previous work [1] which examined the potential of modern machine learning methods to identify possible indicators of dyslexia in handwriting. We hope this system can eventually serve as a preliminary screening tool, which, together with evidence from teachers and parents, can be used to refer the child to a school psychologist for a standard evaluation. Unlike other problems in machine learning, the importance of understanding whether the network is making well-supported decisions cannot be stressed enough—this is not the classification of cats and dogs; these are children's lives. The implications of using a system in the education realm that we do not

AI for Social Good workshop at NeurIPS (2019), Vancouver, Canada.

completely understand are potentially enormous. Thus, we seek to demonstrate the soundness of this technology before we move forward.

2 Related work

2.1 Handwriting recognition

Optical character recognition (OCR) is a well-developed, reliable method for processing handwriting into text. Applications include automated mail sorting [12] and scanning written forms into text, among others. OCR can recognize adult handwriting, including cursive [13], with high accuracy. However, children's handwriting has not been studied to the same extent that adult handwriting has, and it is often messy and poorly constrained, making it hard for these techniques to be applicable.

2.2 Related technical approaches

Many groups have used machine learning to make predictions about whether students have dyslexia, using students' test scores [14], demographic information, survey data [15] (or all three [16]) as features. Others have had success diagnosing dyslexia from computerized classification of brain activity [17, 18] or by tracking eye movement of students [19]. Interactive technology applications to detect dyslexia have emerged in recent years that are geared towards spotting red flags [20, 21, 22]. For example, Rello et al. designed a game to detect dyslexia across languages [23].

None of these approaches use raw handwriting, but handwriting is fairly easy to collect (compared to standardized tests or interviews) and still has a substantial amount of meaning.

2.3 The link between reading and writing

The definition of dyslexia is often misunderstood, so experts have debated whether or not writing problems can signify reading problems. If not, any abnormalities in handwriting would indicate writing or fine motor disorders. However, decades of research in school psychology show that this is not the case; reading is intimately connected to writing, and **children who struggle to learn to read often struggle to write**.

"The writing difficulties of students with dyslexia can be partially attributed to their reading difficulties and can manifest in many ways in their writing, such as poor spelling, poor legibility, lack of diverse vocabulary, poor idea development, and/or lack of organization." [24]. Children with dyslexia tend to form letters in unconventional ways, often omitting spaces between words, extending letters below the line and erasing frequently [24].

Furthermore, teachers are not often aware of these subtle differences. Germano et al. recently showed that while 84.3% of the students with dyslexia in their study met the criteria for a handwriting deficiency, when teachers were asked the same questions about the students, they answered "rarely" most of the time, suggesting they were not aware of the handwriting difficulties of the students [25].

Features of dyslexia in handwriting Sally Shaywitz, an expert in dyslexia detection, found that some of the most robust identifiers of dyslexia include mispronouncing common words, not recognizing rhyming patterns, poor spelling, and messy handwriting [26]. However, not all of these can be identified in handwriting. The most relevant indicators to our analysis are related to poor spelling and ultimately the handwriting itself:

- Ascending or descending handwriting (angled up or down) even on lined paper
- Curvatures and angles of the letters "m", "n", "u" and "v"
- Irregular or no spaces between words, inconsistent sizes of letters or general bad form

Previous research has shown that handwriting from students with dyslexia can indeed be distinguished from handwriting from students without dyslexia [1]. *Can it be done?* is a different question from *Can it be done well?*. Next, we show that the same features that allow a school psychologist to identify dyslexia in a child's handwriting are also evident to the machine learning models we use to detect dyslexia.

3 Methods

3.1 Data

We build on a data set of 56 photos of handwriting samples grades K-6, collected from parents [1].

Pilot school system We partnered with a U.S. school system consisting of 1200 children in grades K-6. We used an online survey to collect parental consent and demographic information from parents. The dyslexia diagnoses were self-reported by parents. Here, we present a preliminary subset of this data as an example, consisting of 100 handwriting samples. See Table 1 for a breakdown by grade.

Grade	1	2	3	4	5	6	Total
Dyslexia	5	3	3	1	10	0	22
No Dyslexia	12	23	11	3	17	2	78
Total	17	26	14	4	27	2	100

Table 1: Frequencies of handwriting samples. The full set will contain all 1200 samples.

3.2 Pre-processing pipeline

Following [1, 27, 28], we clean the images by first cropping them into lines of text, re-scaling those lines to the same size, then generating random patches for each line. These patches can be letters or transitions between letters. See Figure 1 for some of the resulting patches of handwriting.

3.3 Experimentation and results

Using an off-the-shelf convolutional neural network, patches of handwriting (size 113x113 pixels) are used as input and the network classifies the patch as either indicative of dyslexia or not. The data set was split 3:1:1, with 60% training samples, 20% testing samples, and 20% validation samples.

Five-fold cross validation with hyper-parameter tuning yielded **77.6%** accuracy in determining whether a patch of handwriting was written by a second-grade student with dyslexia or not. It is clear that the network used certain features in the handwriting to classify the images; however, we cannot be sure that these features align with meaningful indicators of dyslexia without digging deeper.

4 Analysis

Unfortunately, it is notoriously difficult to understand why a neural network model has made a particular classification decision. We did several rounds of analysis to verify that the network is detecting features of *dyslexia* as opposed to meaningless differences. We first investigated the results visually to understand the qualitative nature of the patches of handwriting that were classified correctly and incorrectly. After forming several hypotheses, we solidified these findings quantitatively as well.

4.1 Visual Observations with TSNE

TSNE [29] is a data visualization technique often used to interpret the high-dimensional output of neural networks. We can observe that points plotted closer together represent patches of handwriting that are more similar than the handwriting represented by points farther apart. We ran TSNE 10 times and chose the result with the lowest KL divergence (see Figure 1).

4.2 Quantifying "Messiness"

From observation, it is fairly clear that samples marked as having dyslexia are not as legible as those without dyslexia, but we wanted to quantify this hypothesis. One of the authors hand-labeled patches of handwriting—5100 of them—in various categories of messiness: (1) illegible (cannot make out the letter or transition), (2) partially illegible (can make out a letter but not sure what it is), and (3) legible (can clearly tell the letter or transition).

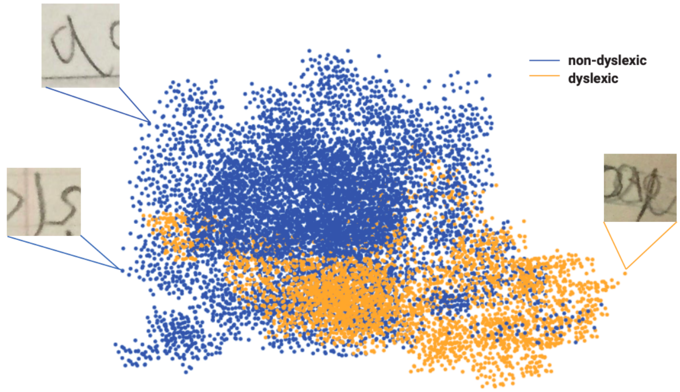

Figure 1: **Dyslexic handwriting tends to be messier.** After plotting patches from second grade, it appears that while there is a lot of overlap, the dyslexic handwriting tends to be less legible.

We labeled patches from the validation set for second grade, then re-validated the network and found that 84% of the patches marked illegible and 60% of the patches marked partially illegible were from samples of children with dyslexia. This seems to confirm that students with dyslexia have objectively "messier" handwriting than their peers, and the network is using "messiness" as an indicator.

5 Dangers of this technology

Now that we have validated this technique, the potential for misuse of this system has risen. While this software was in no way intended to cause harm, there are many possible negative outcomes.

1. **False negatives**: While a false positive would refer a student for testing who may not need it, a false negative has more dire consequences. If that student has dyslexia but has been told they don't, that could prevent them from ever understanding their struggles with reading.
2. **Use outside of the intended use**: What is to stop someone from doing the same thing to try to distinguish varying levels of intelligence in children? It is difficult to know the scope of use that will follow, and even though this project is intended to help, there is always future potential for someone else to do harm.
3. **Bias**: We do not have a completely clean data set. Some of the handwriting we receive labeled as not characteristic of dyslexia could be from a child with an *undiagnosed* reading or writing problem. Additionally, the diversity of our initial pilot test data is limited. Most of the kids with diagnosed dyslexia from the pilot school system are white children with wealthy, college-educated parents. Unfortunately many kids outside of this narrow demographic slice also have dyslexia but do not get detected at the same rates [8].

6 Conclusion and future work

In this paper, we present preliminary work suggesting that the base of our system is solid and the network is detecting more than irrelevant features in the handwriting—it is detecting unique features of language-based learning disabilities. There is a vital need for a system like ours, especially with future sophistication of the methods and additional investigation into the "non-dyslexic" samples.

Our next step is to work closely with a school psychologist to validate the various steps of this system and add other screening measures if necessary. For example, school psychology research has shown [30] that students with dyslexia typically write less when given the same amount of time as their non-dyslexic peers, so we may consider adding a time requirement. Additional research has studied the use of drawings instead of handwriting [31], so we have considered adding that to the screening system as well. Once we have built the system, we will conduct a thorough study and have a school psychologist evaluate each child to determine the true diagnosis.

References

[1] K. Spoon, D. Crandall and K. Siek. Towards Detecting Dyslexia in Children's Handwriting Using Neural Networks. International Conference on Machine Learning AI for Social Good Workshop, 2019.

[2] S.J. Ritchie and T.C. Bates. Enduring Links from Childhood Mathematics and Reading Achievement to Adult Socioeconomic Status. *Psychological Science*, 24(7):1301–1308, 2013.

[3] UN General Assembly. Transforming our world : the 2030 Agenda for Sustainable Development. `https://www.refworld.org/docid/57b6e3e44.html`, 2015.

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

[5] Glaser D. Wilcox D. D. Walsh, K. What Education Schools Aren't Teaching About Reading and What Elementary Teachers Aren't Learning. *National Council on Teacher Quality*, 2006.

[6] 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`, 2017.

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

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

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

[10] National Center for Education Statistics. Back to School Statistics. `https://nces.ed.gov/fastfacts/display.asp?id=372`, 2018.

[11] D. Jones, M. Raby, T. Tolfree, and J. Gross. The Long-Term Costs of Literacy Difficulties. *KPMG Foundation*, 2006.

[12] S.N. Srihari and E.J. Kuebert. Integration of Hand-Written Address Interpretation Technology into the United States Postal Service Remote Computer Reader System. volume 2 of *International Conference on Document Analysis and Recognition*, pages 892–896, Ulm, Germany, 1997.

[13] J. González, I. Salvador, A.H. Toselli, A. Juan, E. Vidal, F. Casacuberta. Offline Recognition of Syntax-Constrained Cursive Handwritten Text. volume 1876 of *Joint IAPR International Workshops on Advances in Pattern Recognition*, pages 143–153, London, U.K., 2000.

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

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

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

[17] F.H. Duffy, M.B. Denckla, P.H. Bartels, G. Sandini, and L.S. Kiessling. Dyslexia: Automated Diagnosis by Computerized Classification of Brain Electrical Activity. *Annals of Neurology*, 7:421–428, 1980.

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

[19] M. De Luca, M. Borrelli, A. Judica, D. Spinelli, and P. Zoccolotti. Reading Words and Psuedowords: An Eye Movement Study of Developmental Dyslexia. *Brain and Language*, 80:617–626, 2002.

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

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

[22] O. Gaggi, G. Galiazzo, C. Palazzi, A. Facoetti, and S. Franceschini. 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)*, pages 1–5, 2012.

[23] L. Rello, K. Williams, A. Ali, N.C. White, and J.P. Bingham. Dytective: Towards Detecting Dyslexia Across Languages Using an Online Game. volume 13 of *Web for All Conference*, Montreal, CA, 2016.

[24] J.B. Hayes P. Bazis S. Cooper M. Herbert, D.M. Kearns. Why Children With Dyslexia Struggle With Writing and How to Help Them. *Language, Speech, and Hearing Services in Schools*, 49:843–863, 2018.

[25] C. Giaconi G.D. Germano and S.A. Capellini. Characterization of Brazilians Students with Dyslexia in Handwriting Proficiency Screening Questionnaire and Handwriting Scale. *Psychology Research*, 6:590–597, 2016.

[26] S.E. Shaywitz. *Overcoming Dyslexia: a New and Complete Science-Based Program for Reading Problems at Any Level*. A.A. Knopf, 2012.

[27] L. Xing and Y. Qiao. DeepWriter: A Multi-stream Deep CNN for Text-Independent Writer Identification. volume 15 of *International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 584–589, Shenzhen, China, 2016.

[28] P. Dwivedi. Handwriting Recognition Using TensorFlow and Keras. `https://towardsdatascience.com/handwriting-recognition-using-tensorflow-and-keras-819b36148fe5`, 2018.

[29] L.J.P. van der Maaten and G.E. Hinton. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[30] V. Connelly E. Sumner and A.L. Barnett. The Influence of Spelling Ability on Handwriting Production: Children With and Without Dyslexia. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 40:1441–1447, 2014.

[31] Q. Rankin and H. Riley. Including Dyslexics: Indicators of Dyslexia in Art Students' Drawings. doi 10.13140/RG.2.1.3159.3682, 2015.