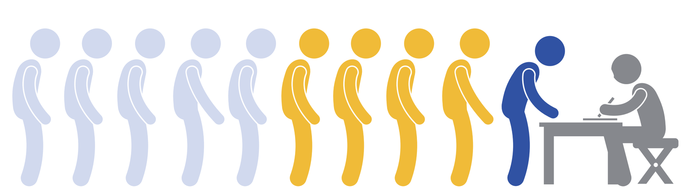
Towards Detecting Dyslexia in Handwriting using Neural Networks

Katie Spoon, David Crandall, Katie Siek, School of Informatics, Computing and Engineering, Bloomington, IN

Introduction

10.2 million
kids in the US with dyslexia will not be diagnosed by the recommended age[2]

- **Literacy** is the best predictor of success later in life. **Dyslexia** is a **learning disability (LD)** characterized by a difficult to read or interpret words, letters, and symbols according to their shapes[1].
- Dyslexia is **not tied to IQ**, which is a common myth. Students with dyslexia perform just as well with **accomodations**. To get these accomodations, students meet with a school psychologist. However, there is a long waitlist and diagnosis is often delayed.



- If students are placed in the waitlist by the end of 2nd grade, they improve their chances of graduating from high school[2]. However, **teachers usually don't have the training** to detect LDs[3] and often detection comes too late.

How can machine learning help?

Machine learning (ML) can **recognize patterns that we don't see** (or don't see easily). We believe this problem is ideal for ML due to the lack of an organized dyslexia detection process and human power to carry out this detection, along with numerous subtleties in the definitions of language-based LDs.

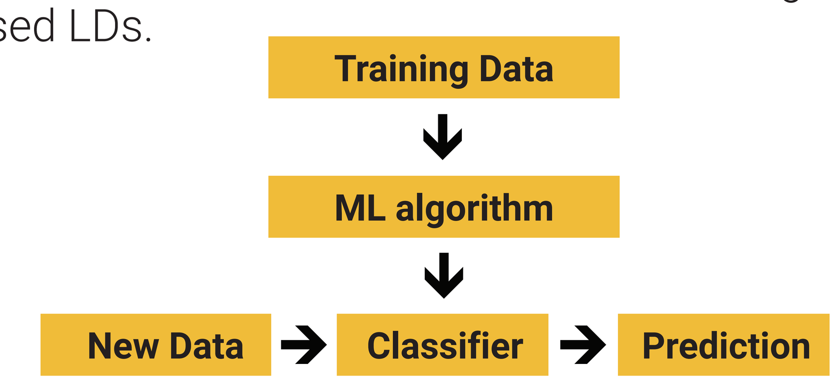


Fig 1. **Machine learning in a nutshell.** The algorithm learns using training data, then new data is given to the classifier, which uses the algorithm to predict the classification of the new information.

Our Approach

Goal: Create a tool using a subfield of ML, deep learning, to **identify characteristics of dyslexic handwriting** and place students into the diagnostic queue by the end of 2nd grade

Neural Network Architecture & Potential Feature Identification

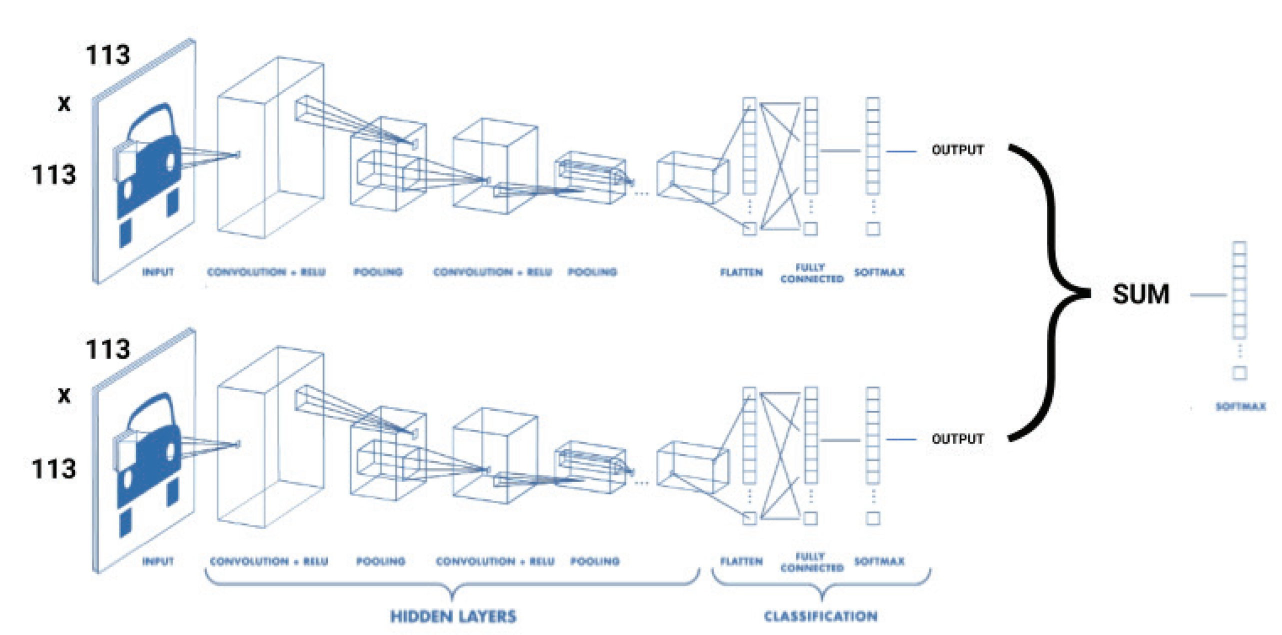


Fig 2. **Our model.** This model[4] was used for a different problem. It takes in **two patches, one from dyslexia and one without**. It has 5 convolutional layers, 3 max-pooling layers, and 2 fully-connected layers. The convolutional and fully-connected layers are **shared** between the two streams. The results of the streams from both patches are added to get the final result. This approach is used to help **discover the polarizing features** between the two patches.

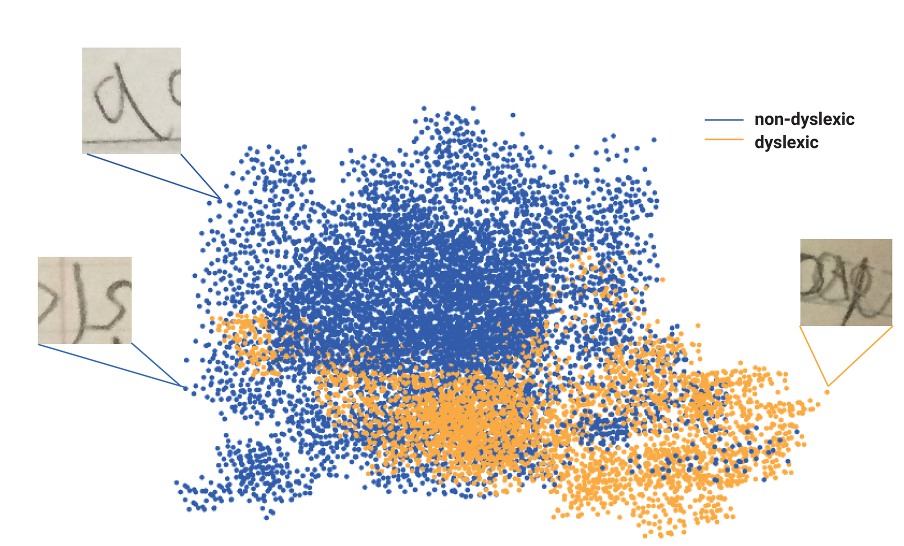
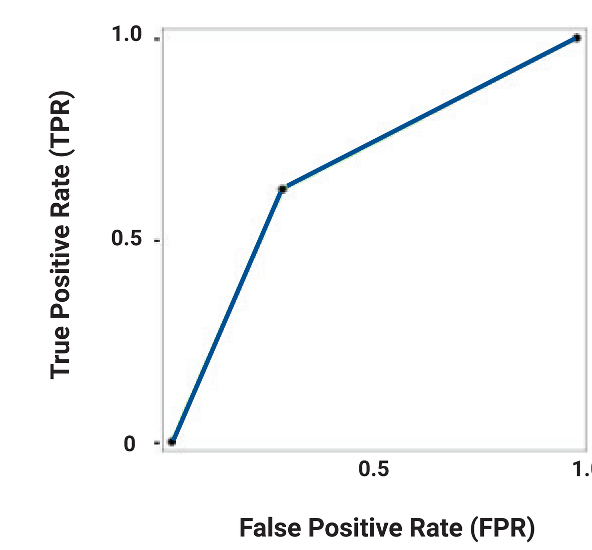


Fig 3. **Dyslexic handwriting tends to be messier.** TSNE is an ML **visualization technique** that maps a high-dimensional feature vector (in our case, from the last layer of a neural network) into 2D space. **Points closer together are more similar**. After plotting a sample of patches from 2nd grade, it appears that while there is a lot of overlap, the dyslexic handwriting tends to be messier or harder to identify. Outliers were removed for clarity.

Implications

Higher Early Detection Rate Than Teachers



68%
accuracy in detection, up from 2%

Fig 4. **Area Under the ROC Curve (AUC)** for the classification of dyslexic versus non-dyslexic students in second grade was **68%**. Currently, only 2% of students are detected by this point by teachers[2].

Why does it matter?

- Children who struggle to read in third grade are **four times more likely** to drop out of high school[5], and will struggle with reading, math and self-esteem for the rest of their lives. The UK estimates that the country spends **one billion pounds annually**[6] as a result of undianosed dyslexia - these undiagnosed students are more likely to struggle with employment, spend time in prison and have health issues, **straining the economy.**
- The current detection process is biased and can over-detect or under-detect certain groups of people[2]. Some teachers and parents are more attentive than others, causing additional bias. Our method may **detect more students**, and fairly.

Future Work

- We will improve the model for detecting dyslexia after we **collect more data**, build a **front-end application** for teachers and parents, and utilize more data visualization techniques to better understand which features the network uses to make decisions.
- We would also like to eventually **expand the system** to detect different types of LDs and to monitor English Language Learners, as the tool is not language-specific.

Contact
Katie Spoon: kspoon@iu.edu

1. Learning Disabilities Association of America (LDA)
2. National Center for Learning Disabilities (NCLD)
3. National Council on Teacher Quality (NCTQ)
4. Priya Dwivedi, English Deep Writer, (2018), Github repository, https://github.com/priya-dwivedi/Deep-Learning/tree/master/handwriting_recognition
5. Hernandez, D. J. (2012). Double jeopardy: How third-grade reading skills and poverty influence high school graduation. Baltimore, MD: The Annie E. Casey Foundation.
6. British Dyslexia Association