CLASSIFYING DYSLEXIA BASED ON HANDWRITING

ATЯAIN

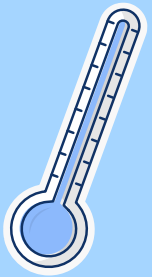
Amishi Gupta, Tyler Westland, Ritvik Irigireddy, Angel Li, Isha Prem, & Neta Shubin

TABLE OF CONTENTS

01

Introduction

What Is Dyslexia?



Solution

Issue

Dyslexia is a general term for disorders that affect your writing, speaking, reading, and spelling skills, impacting a student's ability to read and write.
This is why it is important that it is diagnosed and treated early.

Strategy

Generate and optimize a machine learning model that can classify if an individual is likely to have dyslexia in order to initiate further potential consultations.

Aim

To create an application using our machine learning model to predict if an individual has dyslexia based on their handwriting.

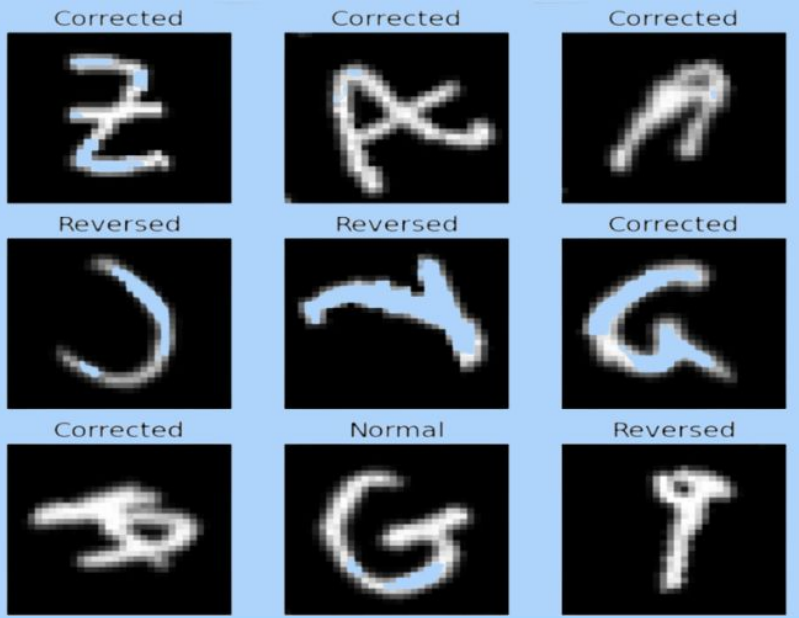
02

Dataset

What are we working with?

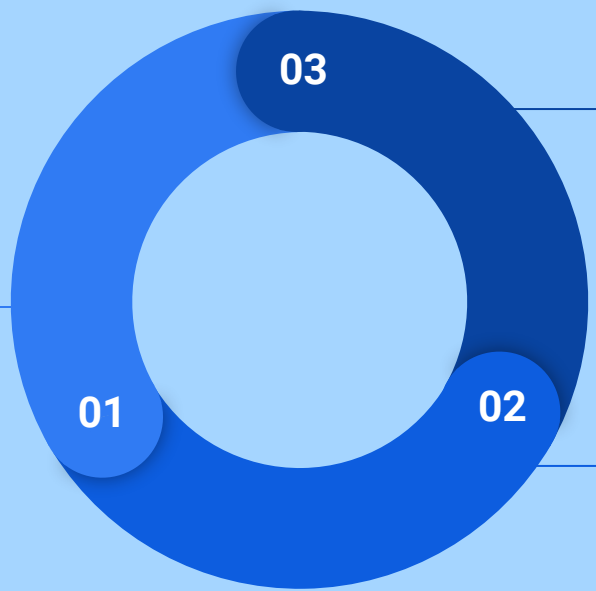
Our Data

https://www.kaggle.com/datasets/drizasazanitaisa/dyslexia-handwriting-dataset

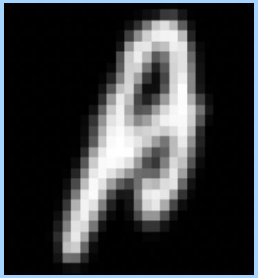


Classes	
Corrected	Characterized as messy handwriting
Normal	No issues with legibility and orientation of letters
Reversal	Letters are in the opposite orientation as normal letters should be and are therefore indicating dyslexia

Preprocessing the Data

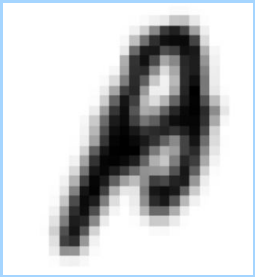


Original Data

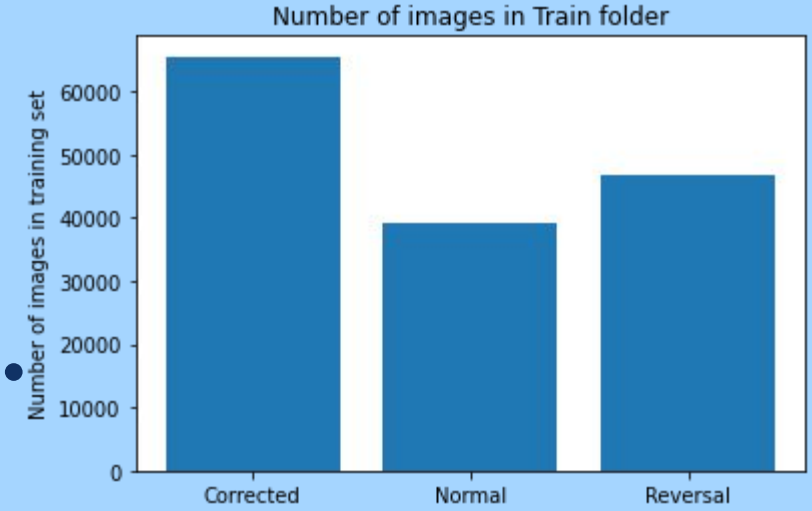


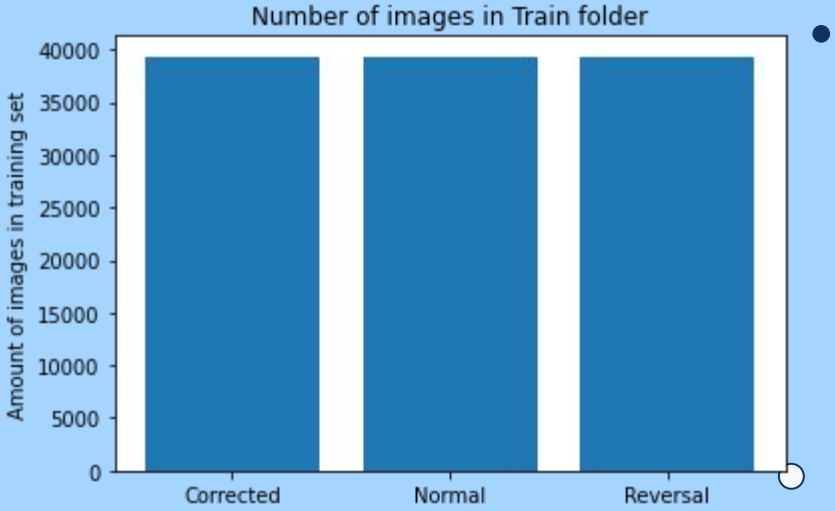
Rotated Data

Preprocessed Data



Undersampling The Data

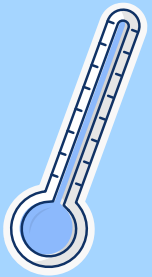




03

ML Models

What tools can we use?



How Are Models Evaluated?

Accuracy

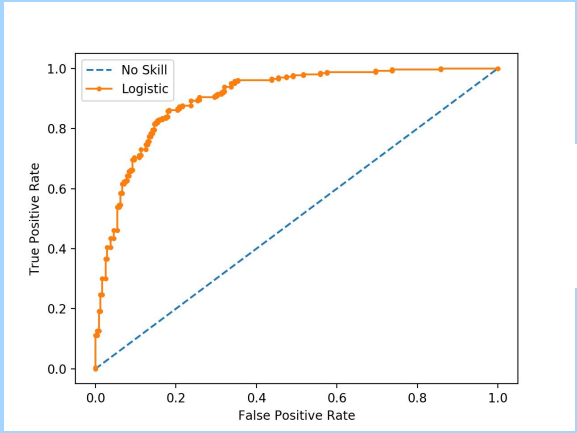
- Percentage of predictions that were correct on our test set.

Confusion Matrix

		Predicted Class	
		0	**1**
True Class	**0**	True Negative	False Positive
	1	False Negative	True Positive

Receiver Operating Characteristic (ROC) Curve

- The greater the area under the curve (AUC), the better the model performs



$$\text{True Positive Rate (TPR)} = \frac{TP}{TP + FN}$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN}$$

Models

K-Means

Our first image classifier

Random Forest

Used to predict accuracy and classification

CNN

A type of neural network with convolutional layers

LeNet-5

A type of CNN model for handwritten digit recognition

Inception V3

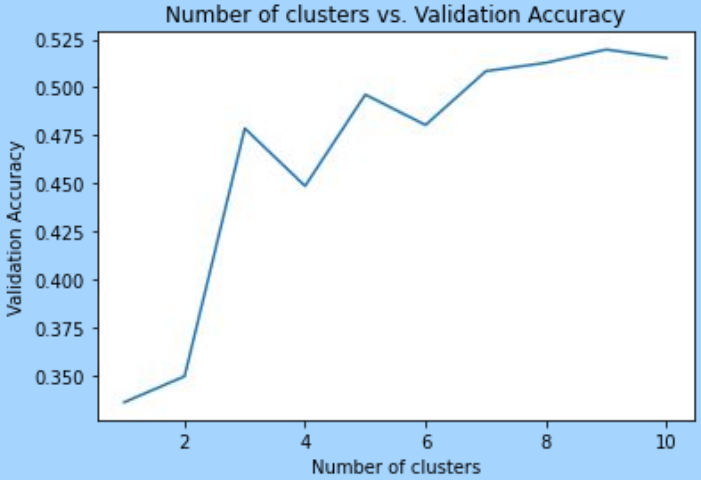
Uses the pre-trained convolutional layers of a model

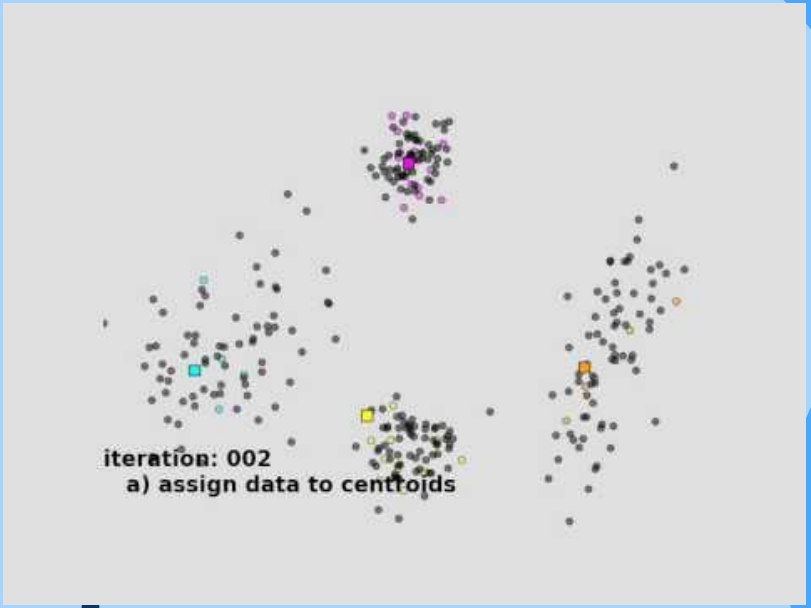
VGG16

Another pre-trained model that performed very well with our dataset

K-Means

Poor accuracy no matter the number of clusters because the images in our dataset aren't easily clusterable.



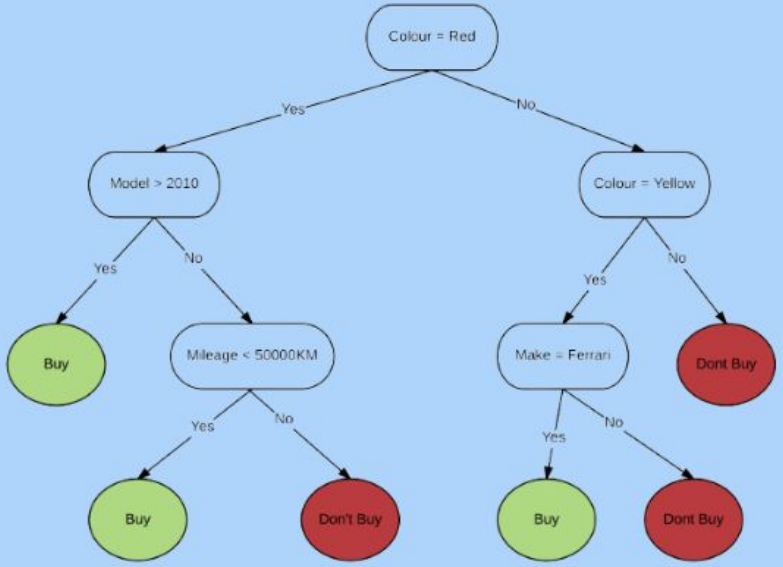


Random Forest

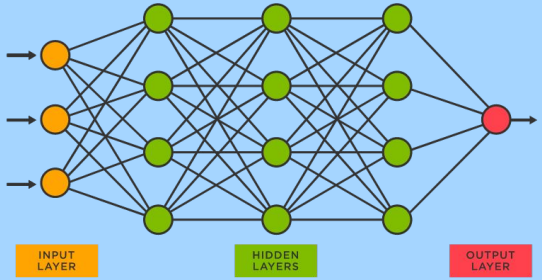
THE CONCEPT

A combination of multiple decision trees that extracts the most important features.

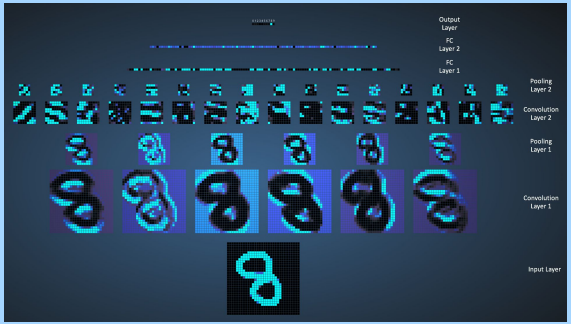
RandomizedSearchCV



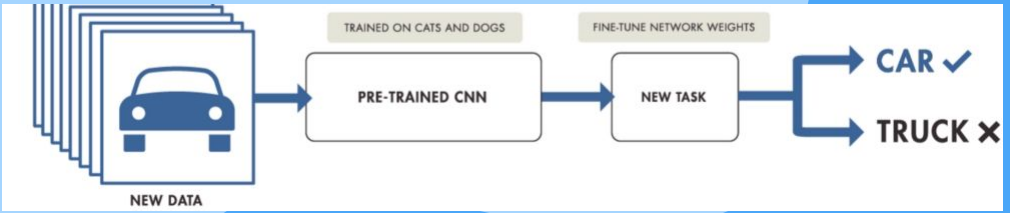
Neural Network



Convolutional Neural Network (CNN)



Transfer Learning



CNN: LeNet-5

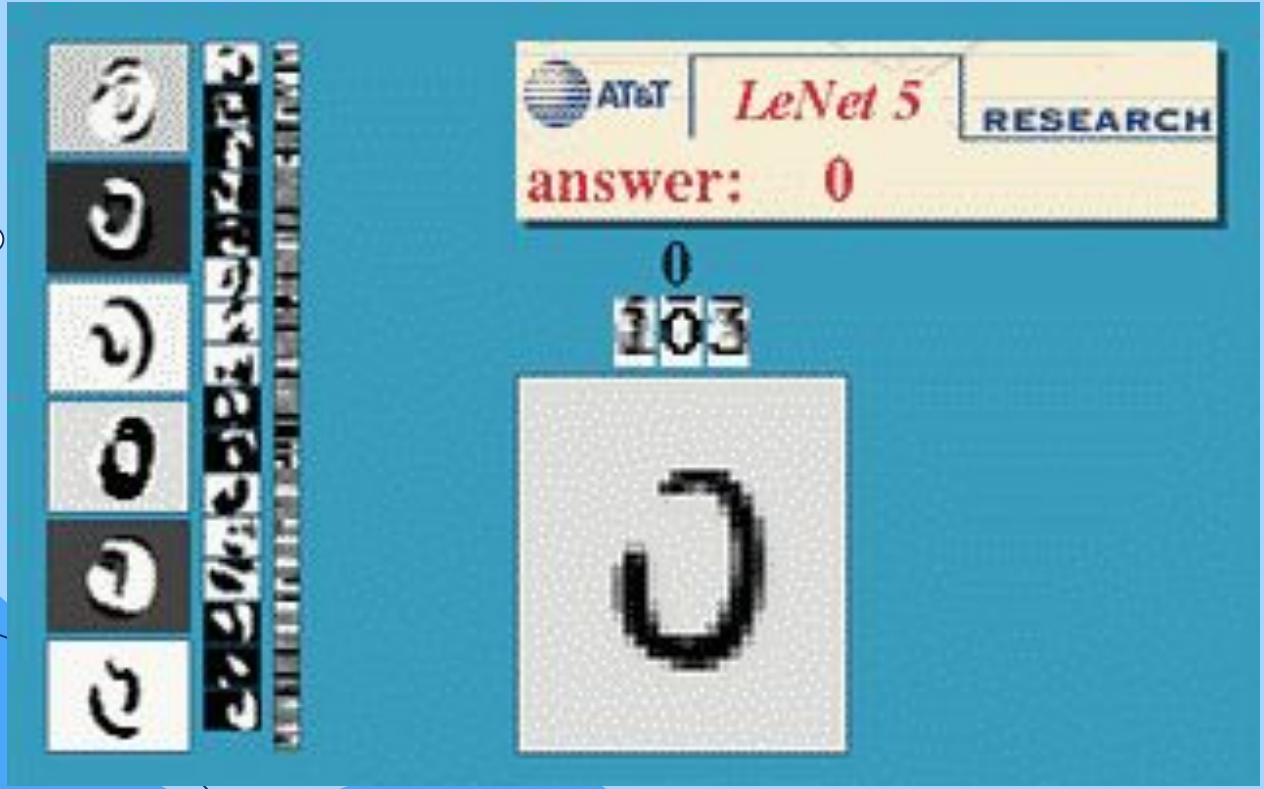


Image size: 32 x 32
2 Convolutional Layers
2 Average pooling Layers
2 Dense Layers
RMSprop Optimizer
Learning Rate: 0.001

Transfer Learning: InceptionV3

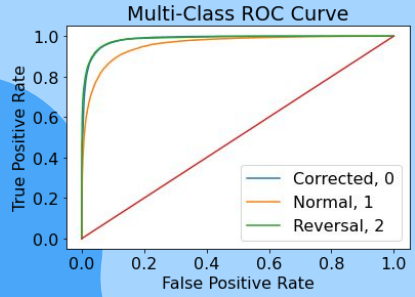
1. **Image Size**
   75 x 75
2. **Layers**
   1 Global Average Layer, 2 Prediction Layers (One has 64 units and the other had 3 units both has relu activation functions)
3. **Loss Function and Optimizer**
   Categorical_Crossentropy and Adam Optimizer
4. **Learning Rate**
   0.0001

The Best Model - VGG16 Model

Avoids Overfitting

The training accuracy (87.8%) and validation accuracy (89.4%) are similar indicating minimum overfitting

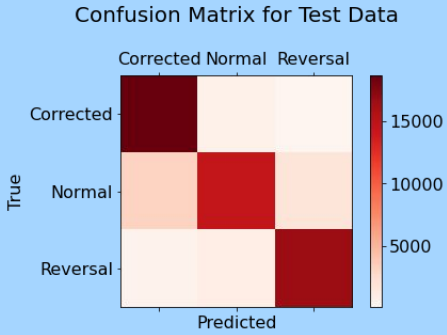
ROC Curve



Uses Preprocessed Data

Data utilized is realistic after being preprocessed

Confusion Matrix



04

Collaboration

BOSTON NORTH Children's Dyslexia Center

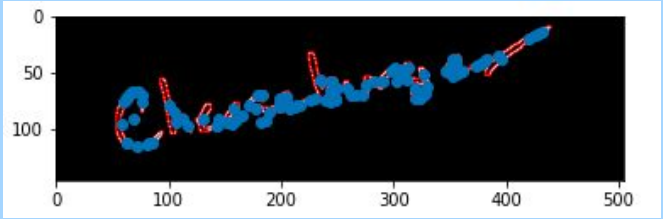
How Collaborating Helped Us?

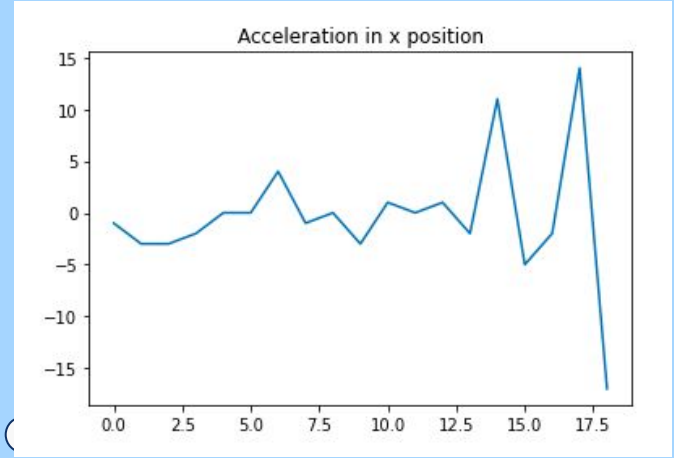
- Letter reversal can occur in younger individuals without dyslexia. Not all individuals with dyslexia perform letter reversals
- "Although students with dyslexia produced a higher proportion of reversal errors than those without dyslexia in bother letter writing and letter naming" Diane Baum, Director of Children's Dyslexia Centers - Boston North.

- Adding a disclaimer for user age (the user must be above the age of 7 in order for the results to be more accurate)
- Adding an algorithm to determine if there is an abnormal spacing between letters and words (another indicator of dyslexia).

Abnormal Spacing Algorithm

- Algorithm detects abnormal spacing between letters in a word
  - OpenCV
  - Contour Centers
- Would like to use a machine learning dataset to improve upon this algorithm in the future



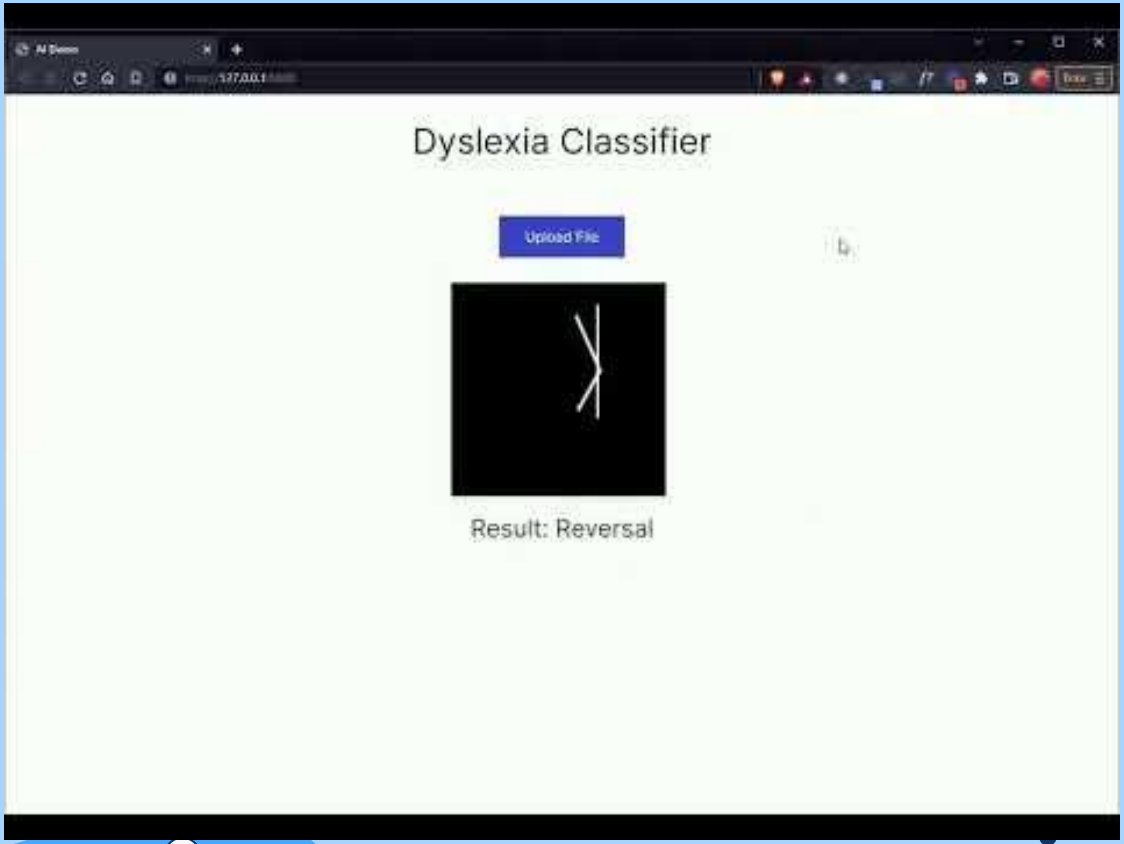


05

Web App

Demo of Our Application
Built with Flask

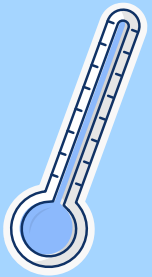
Web Application Demo Video



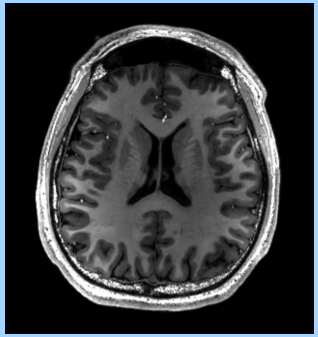
06

Future

Improvements &
Implications



Further Implications



01

More Features

Including more detections for features in dyslexia handwriting (ex: abnormal mix of upper and lower case)

02

Collaboration

Continue working with and learning from Children's Dyslexia Centers

03

Datasets

Supplemental MRI scans

04

Mobile App

Provide an application with a friendlier and more versatile user interface

Thank You For Listening!

— Angela, Amishi, Isha, Neta, Ritvik, & Tyler

Email TylerWestland@gmail.com regarding any questions!