DATA+AI SUMMIT 2022

Machine Learning Models to Aid Autism Diagnoses

Using AI for interpretable diagnoses on sparse data

Anish Lakkapragada

Researcher, Stanford University and Lynbrook High School

ORGANIZED BY databricks

Who am I?

Thanks for asking!

- Rising Junior at Lynbrook High School, San Jose, CA
- Professional ~~copy and paster~~ coder
- r/ProgrammerHumor Junkie
- Researcher at Wall Lab, Stanford University

Agenda

Where we are headed

- ML Workflow
  - Introduction
  - Mission
  - Data Processing
  - Model Training
  - Results
- Demo
- Q&A Session

Introduction

Let's get started

- Worked at the Wall Lab ('21–'22)
- Developed a novel ML model(s) for classifying a key indicator of autism in videos for a faster diagnosis
- Demonstrates that ML is feasible for *explainable* diagnoses in healthcare
- Paper is published in JMIR Biomedical Engineering
- Preprint here: https://arxiv.org/abs/2108.07917

Create an ML model to detect hand flapping from crowdsourced videos to aid in an autism diagnosis.

Building a model

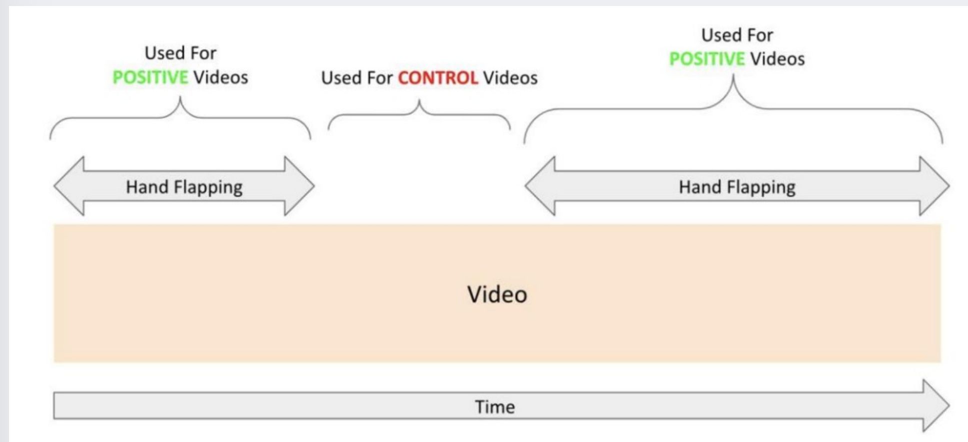
The bare necessities of creating an ML model

- Needed to collect relevant data
- Design a model to train
- Train that model
- Design a fair evaluation technique
- Evaluate that model
- Reiterate

Data Collection

Finally, we're talking data at a data conference!

- Needed videos of hand flapping (and without hand flapping)
- Self Stimulatory Behavior Dataset (SSBD) is the only publicly available dataset with this information
- ≈25 YouTube videos depicting hand flapping stored in XML Files



Data Cleaning

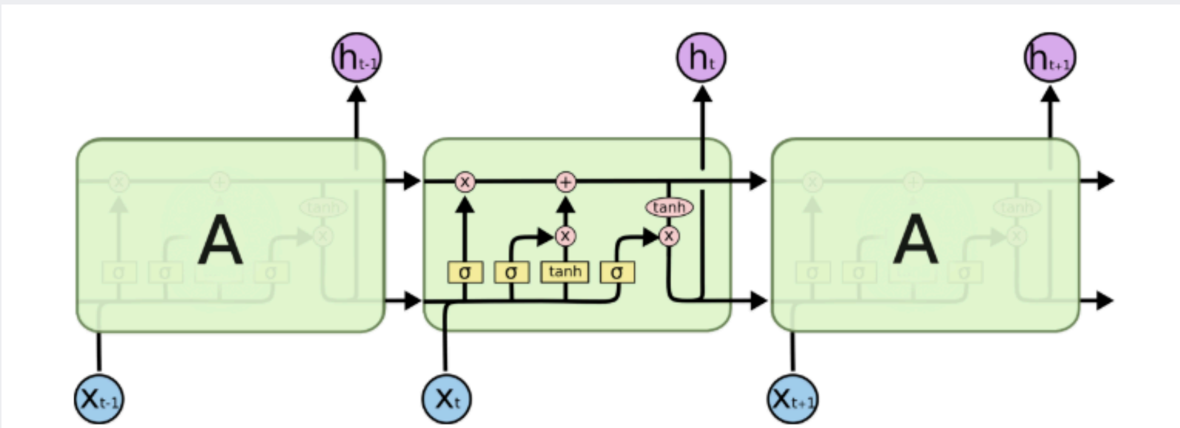
"Bad Data, Bad Data Everywhere"

- Shakiness + Excessive Motion in SSBD
- The application would be on a phone, so videos will be shaky
- Manually cleaned it up

Designing The Model

Long-Short Term Memory Primer

Sorry, it's not a transformer…



The repeating module in an LSTM contains four interacting layers.

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

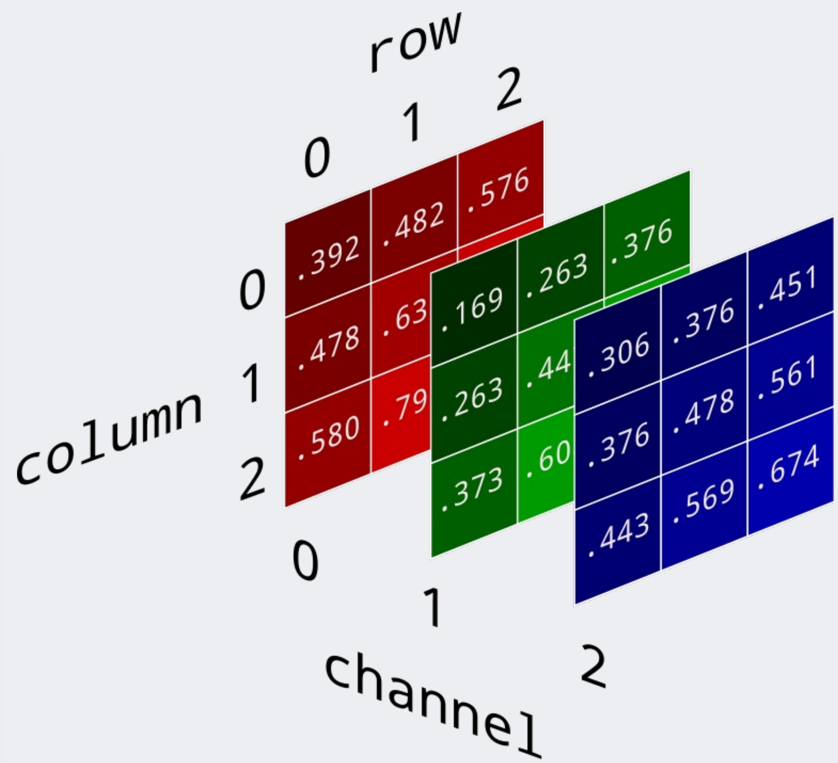
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Who even cares about this
when TensorFlow exists

Images are Huge

And machines don't "see" them

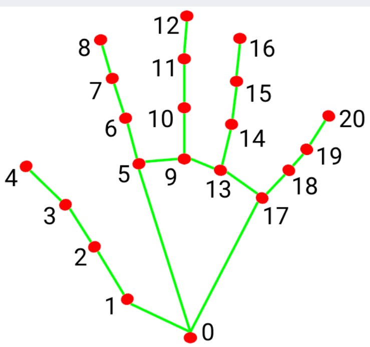
- Usually are around 10^6 floats
- Cannot squish images into vectors and feed them in
- Need to reduce dimensionality



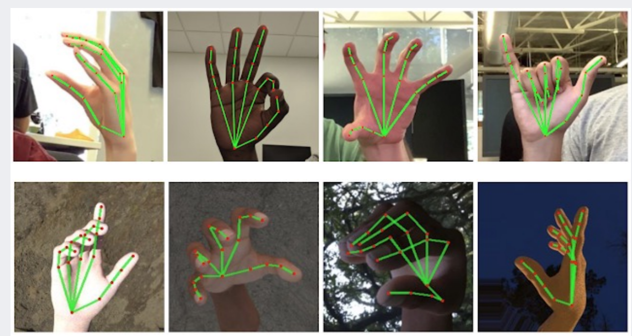
Approach 1: MediaPipe

Gotta start somewhere

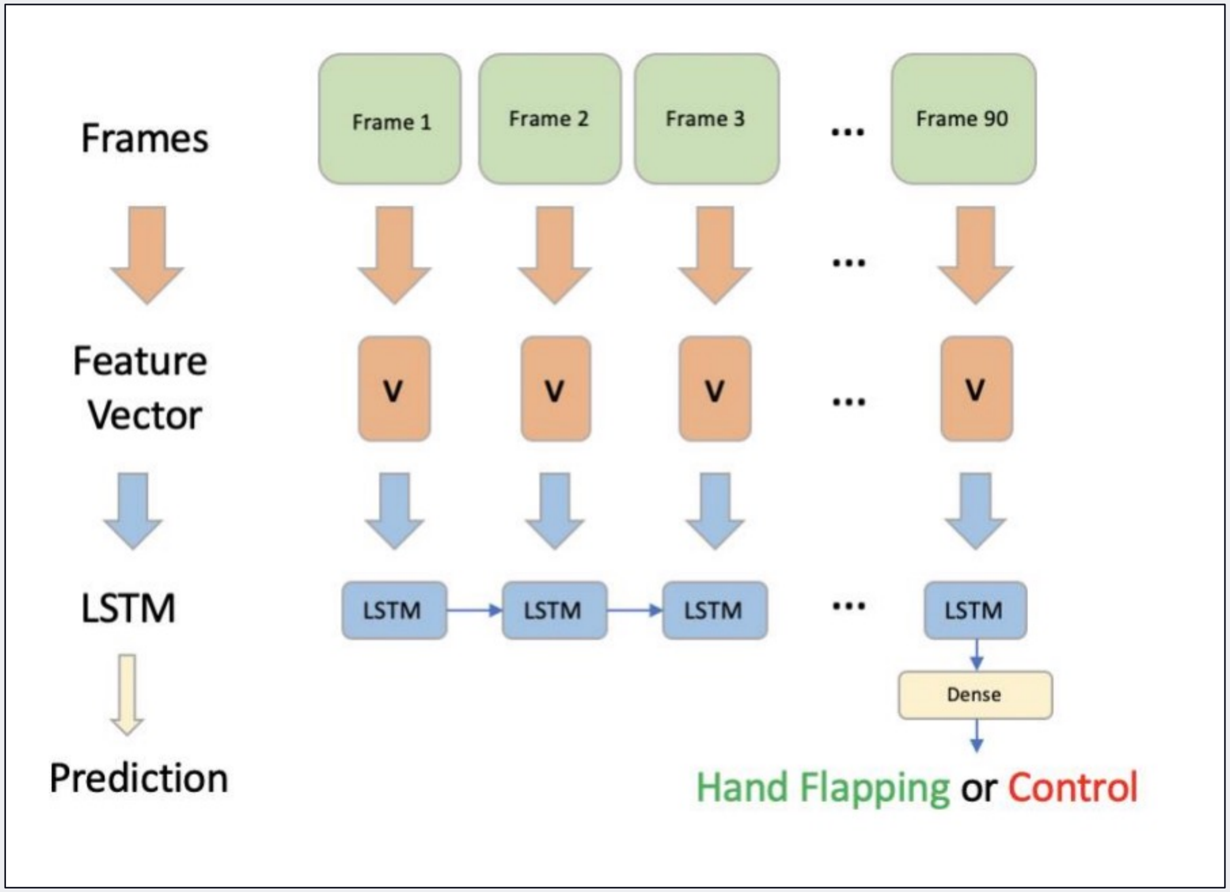
- Made by Google for Pose Estimation and Landmark Detection
- Landmarks = Key Points
- Used Hand Detection => returns the (x, y, z) coordinates for 21 landmarks



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP



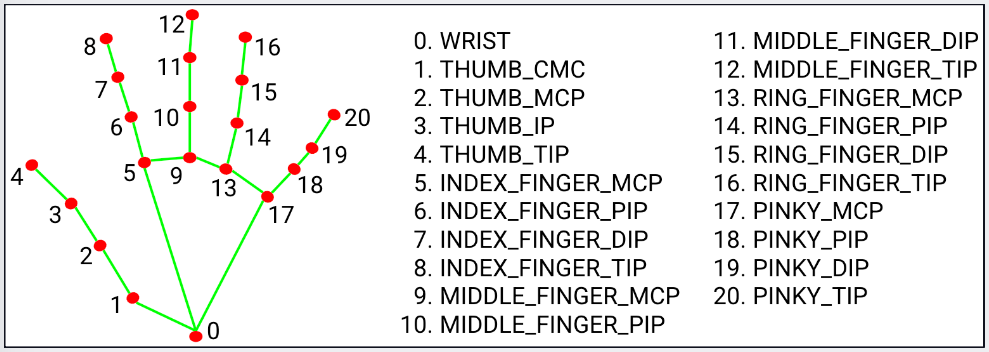
Quick Sketch



Variants of Approach One

Just different takes on the same idea

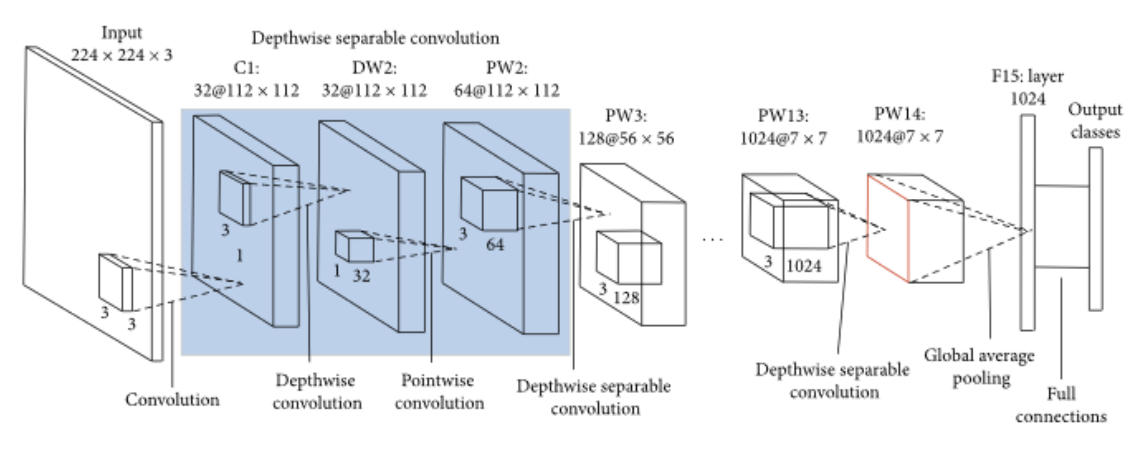
- All Landmarks: Use all 21 Landmarks on Each Hand
- Six Landmarks: Only use finger tips
- Mean Landmark: Take the mean location
- Single Landmark: Only use 1 landmark



Approach 2: MobileNet V2 Pretraining

Just use

- Used MobileNet V2's Convolutional Layers Pretrained on ImageNet
- Fed the extracted vector from these conv layers into the LSTM

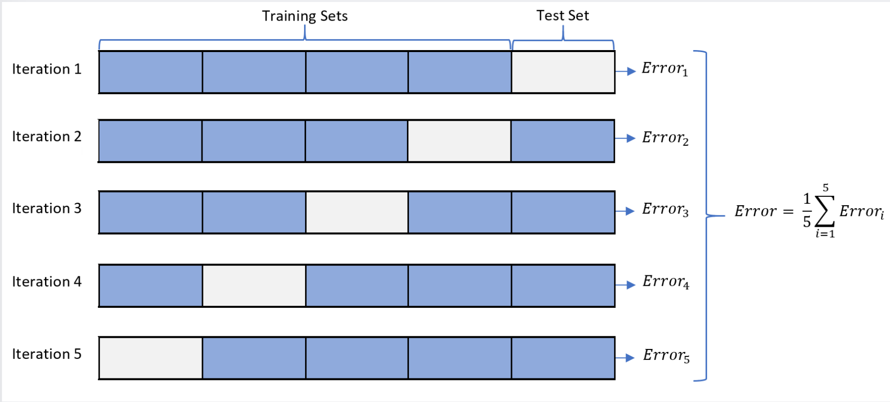


Evaluation Technique

Fair Evaluation: 5-fold Cross Validation 100x

`No cherry-picking!`

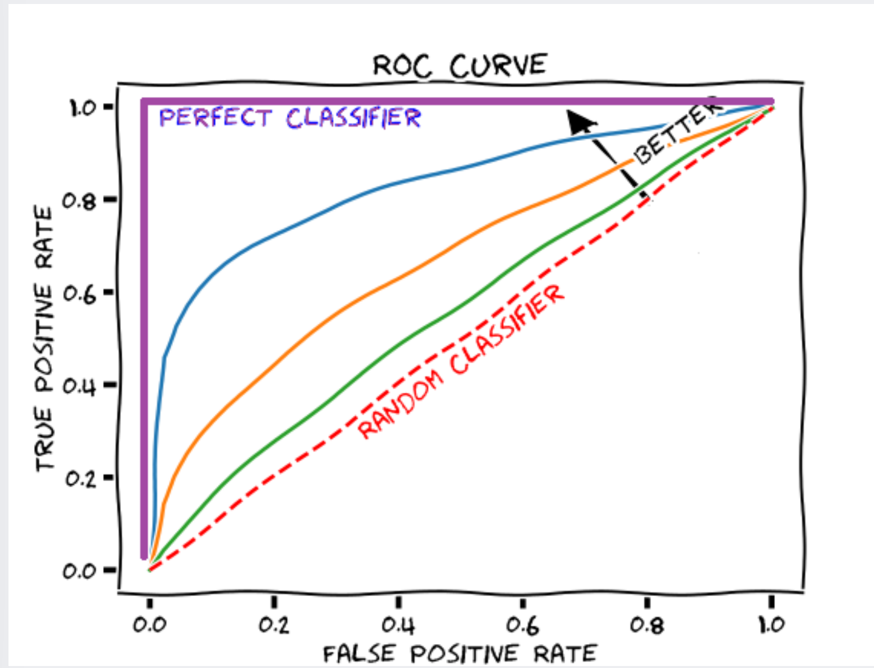
- Created 100 different datasets of 5 folds, and ran 5-fold cross validation on each of them
  - For all approaches (all, six, one, mean landmarks + mobile net)
- Needed for an objective measurement



Metrics

Can't manage what you can't measure

- ROC Curve (*receiver operating characteristics*)
- Accuracy, precision, recall, and F1
- Tracked both training + testing



Results

What I've been stalling all along

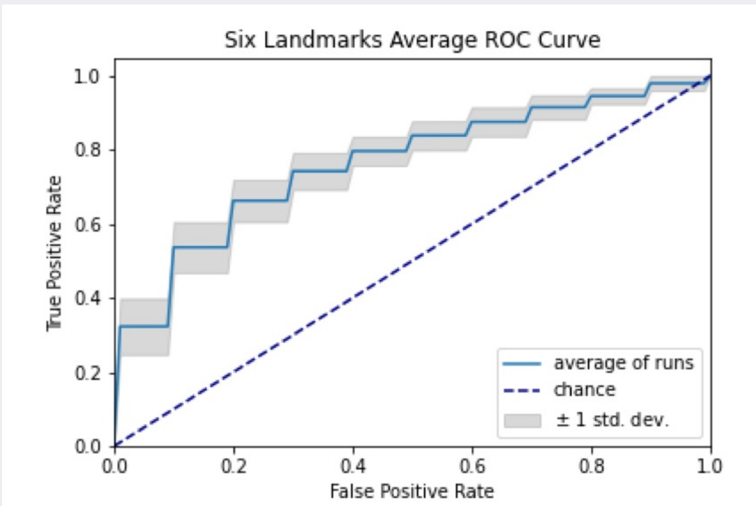
Approach 1: MediaPipe

- Got in the high 69-70%
- Could not overfit to 100% accuracy
- More Fluctuations between variant approaches
- Overall:
  - 1) Six Landmarks
  - 2) All Landmarks
  - 3) Mean Landmark
  - 4) One Landmark

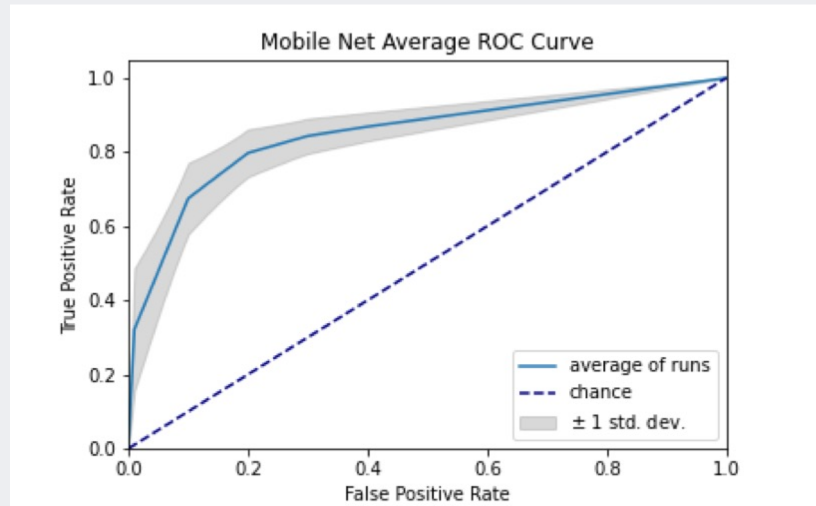
Approach 2: MobileNet V2 Pretrained

- Hovered in the 85% accuracy range
- Had capacity to overfit to 100% accuracy
- Overall a much more accurate model

More Comparisons



Run Type	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Training	76.8 ± 1.95	78.7 ± 2.9	74.7 ± 3.5	76.2 ± 2.1
Testing	69.55 ± 2.7	71.7 ± 3.5	67.5 ± 5.5	68.3 ± 3.6



Run Type	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Training	97.7 ± 1.0	99.5 ± 0.0	95.9 ± 1.7	97.6 ± 1.0
Testing	85.0 ± 3.14	89.6 ± 4.3	80.4 ± 6.0	84.0 ± 3.7

Code Demo

https://github.com/anish-lakkapragada/Hand-Classification-For-Autism-Diagnosis

Q&A Session!

DATA+AI SUMMIT 2022

Thank you!

Anish Lakkapragada
Wall Lab @ Stanford University