A black and white photograph showing four hands in various sign language gestures against a black background. The top left hand is in a fist-like position. The top right hand is open with fingers spread. The bottom left hand is open with fingers spread. The bottom right hand is pointing forward with the index finger.

# **AUTOMATE SIGN LANGUAGE TRANSLATION INTO TEXT**

October, 2022  
FourthBrain, Inc.

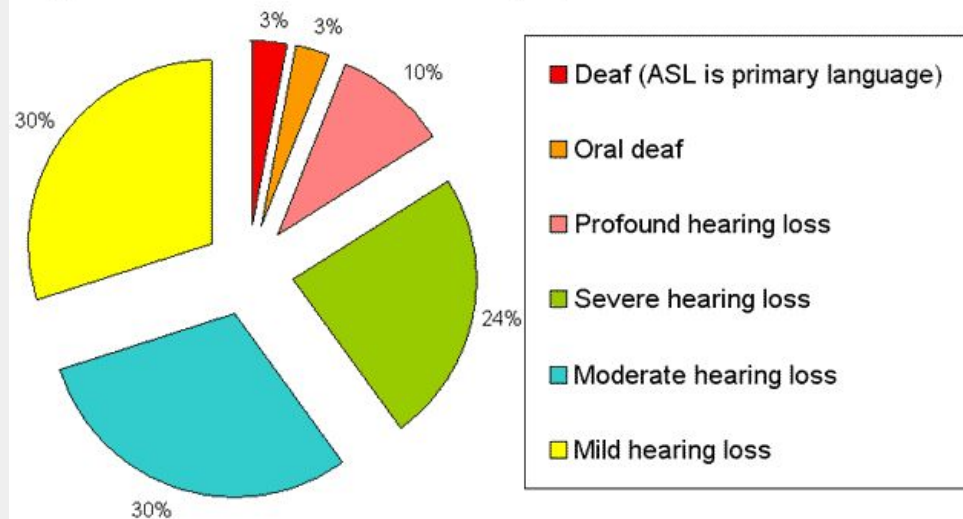
Sheilah K.

# AGENDA

1. PROBLEM
2. SOLUTION
3. DATA + MODEL
4. DEMO
5. SYSTEM DESIGN
6. ETHICAL CONSIDERATIONS
7. FUTURE WORK
8. Q&A

# OVERVIEW

## Approximate Deaf and Hard of Hearing Population in the United States



### Sources:

- World Health Organization
- US Census Bureau
- National Institute on Deafness and Other Communication Disorders

## Percentage of Individuals with Hearing Loss by Age & Severity

MILD

MODERATE+

**38.2 Million (14.3%)** Americans aged 12 years or older have hearing loss

0.2%

12 - 19

0.4%

20 - 29

1.6%

30 - 39

6.5%

40 - 49

13.3%

50 - 59

26.8%

60 - 69

54.6%

70 - 79

81.5%

≥ 80

JOHNS HOPKINS  
BLOOMSBURG SCHOOL  
OF PUBLIC HEALTH

Cochlear Center for  
Hearing and Public Health

jhucochlearcenter.org

Goman & Lin. (2016). Prevalence of hearing loss by severity in the United States. American Journal of Public Health, 106(10), 1820-1822.

**5%**

According to the World Health Organization, there are 466 million deaf people in the world (432 million adults and 34 million children)

**90%**

Although it is believed that the majority of Deaf people learn Sign Language from their Deaf mothers and fathers, the fact is that 90% of Deaf people come from hearing families and learn Sign Language outside the family



# 01

# PROBLEM

# BUSINESS VALUE: OPPORTUNITY

- **Large addressable market:**

- 14 % of the US population aged 12 and above, have reported some form of hearing loss in both ears (NIDCD surveys)
- Roughly 1 million are functionally deaf — or unable to hear normal conversation even when using a hearing aid

- **Inefficient existing solutions:**

- Hindrance to visual communication– background noise, lighting, pace of conversation, number of speakers, accents, facial hair
- Cost– hiring translators, auditory-enhancing equipment

- **Lack of autonomy due to socio-economic challenges:**

- Unemployment or underemployment
- Education gaps
- Delayed healthcare access
- Social isolation– communication barriers



# 02

## SOLUTION

- Deep-learning based tool that can automatically translate American sign language gestures into English text
- “real -time” usage for a more accessible and inclusive world through AI –comes in handy during emergencies, doctor appointments, job interviews etc.
- **Solution 1: Real-time prediction using webcam**
- **Solution 2: Upload photo via FastAPI web application for prediction**
- Metrics : accuracy and quick inference

## Data processing :

- 2, 520 RGB samples - hand gestures
- 70 samples for each of the 36 classes (A-Z, 0-9)
- open- sourced from kaggle
- in-place augmentation and scaling in batches

# 03

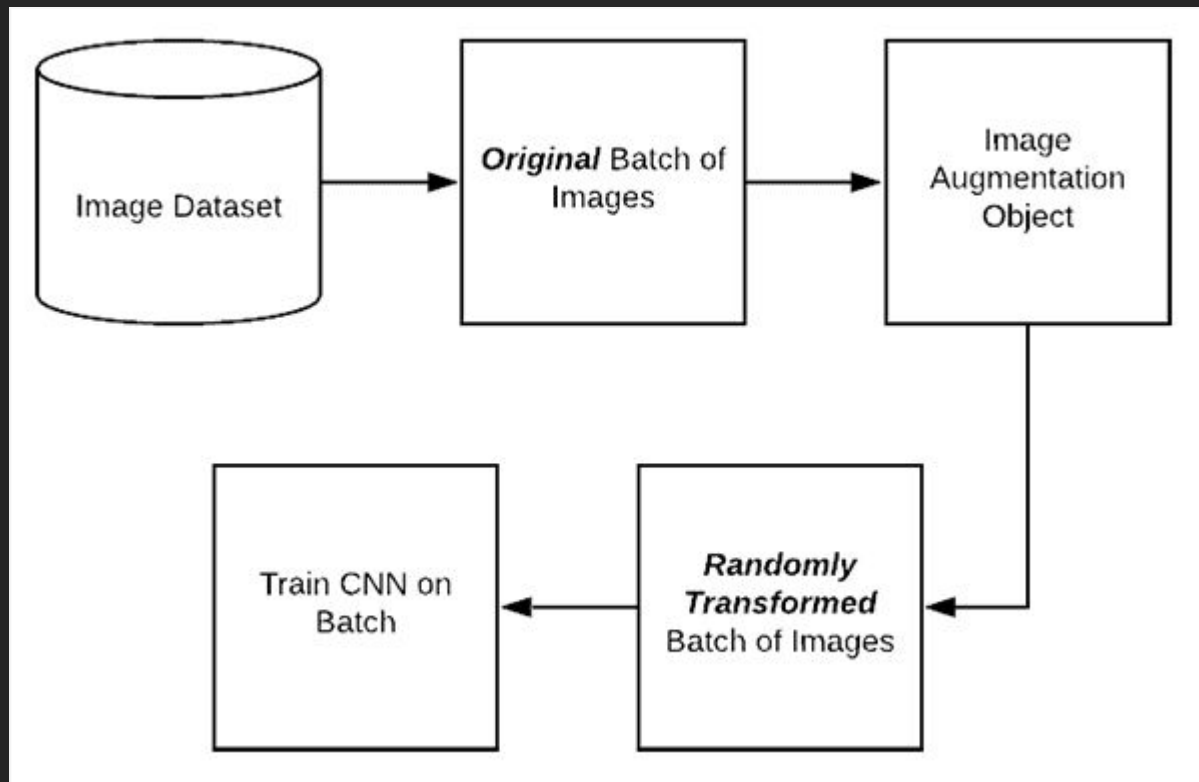
## Model-training :

- Transfer-learning with pre-trained CNN models:
  - vgg16
  - Mobilenet
- Very little fine-tuning:
  - Frozen weights of most layers

# DATA + MODEL



# Data flow



## Keras

ImageDataGenerator

Preprocess\_input function

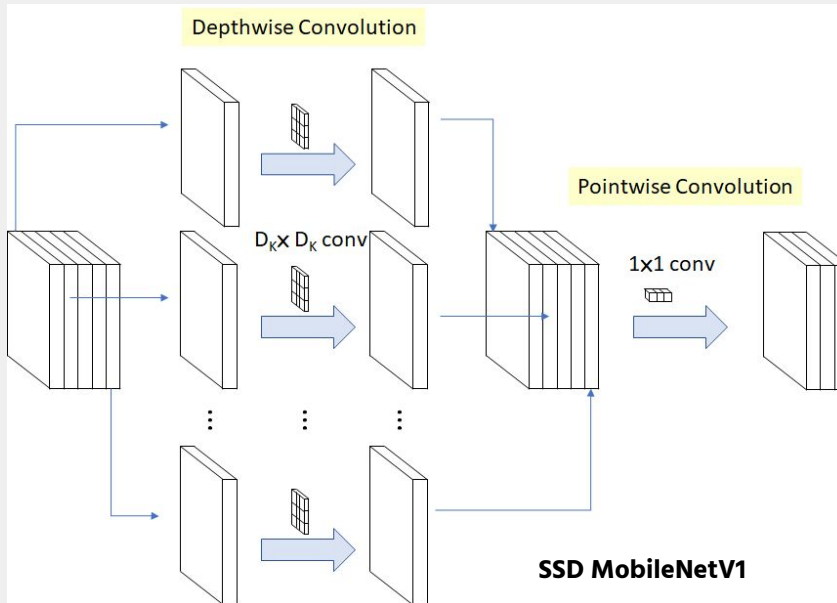
Normalization

In-place augmentation

Resize Input shape

Batch size

# VGG-16



Model	Accuracy
<b>MobileNetV1</b>	0.88
VGG16	0.20

Conclusion: MobileNet is 32X smaller than VGG16 yet is 4X faster, which means it's efficient

# LIVE DEMO

04

A minimalist illustration of a computer monitor. The screen is a solid dark gray rectangle. The base of the monitor is a simple, light gray stand with a flat rectangular foot. The text 'Web API' is centered on the screen in a white, sans-serif font.

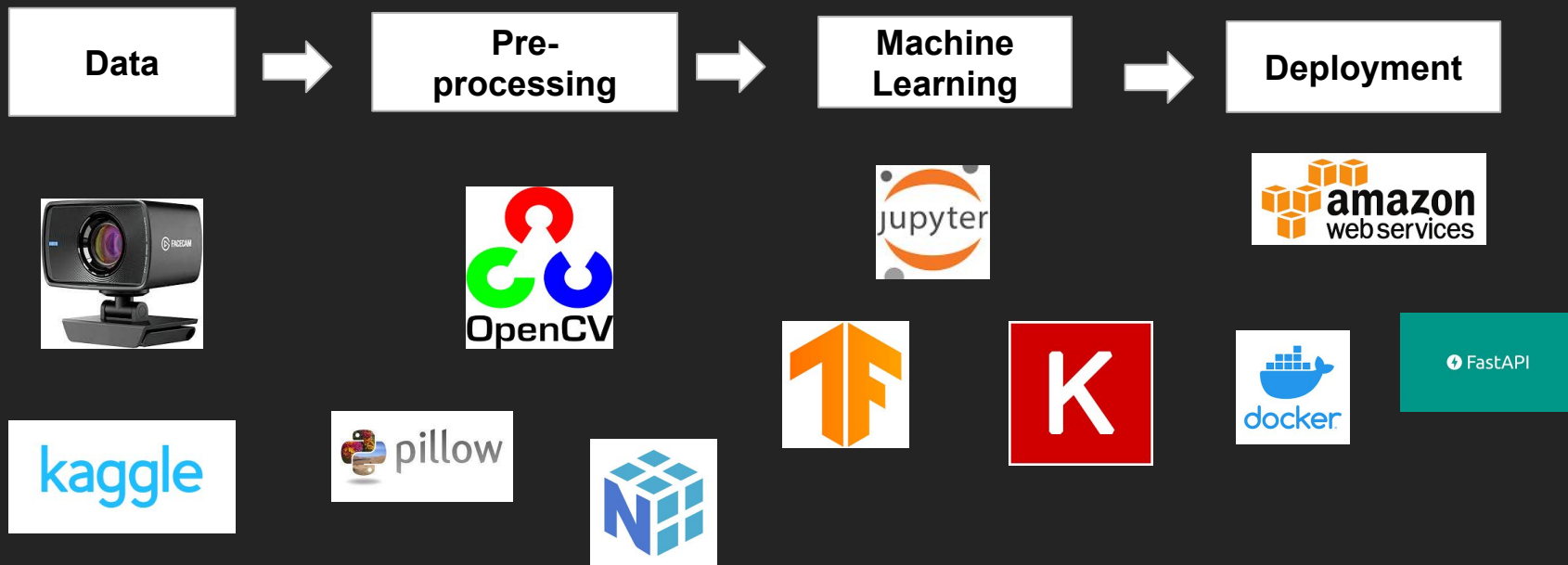
Web API



# 05

## SYSTEM DESIGN

# ML Stack



# 06

- Racial bias: based on skin color and hand features, include images from diverse racial backgrounds
  - Confirmation bias : data is well-structured and curated for image classification
- 
- Inaccurate translations: compromised communication and understanding for deaf users
  - Unknown origins: data might have violated ethical consent when collecting the images

## ETHICAL CONSIDERATIONS

# 07

## FUTURE WORK

- Train on more models/ Extensive hyperparameter tuning to improve accuracy (VGG16)
- Can this model be used to translate gestures into complete sentences?
- Incorporate explainability component
- Deploy real-time system in AWS EC2 instance

**QUESTIONS**