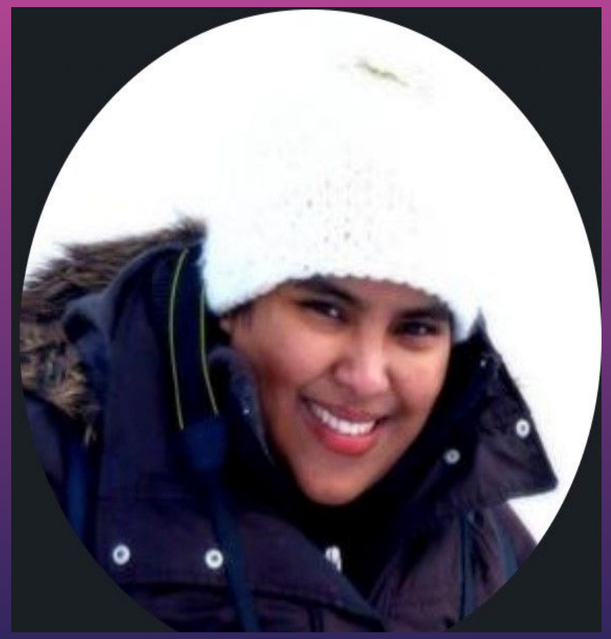
signlanguage.io

Building Inclusive Communities for All

Deepika Maddali, Srila Maiti

The Team

Deepika Maddali



Srila Maiti

Motivation

Blindness cuts us off from things, but deafness cuts us off from people.

-Helen Keller , American deaf-blind educator

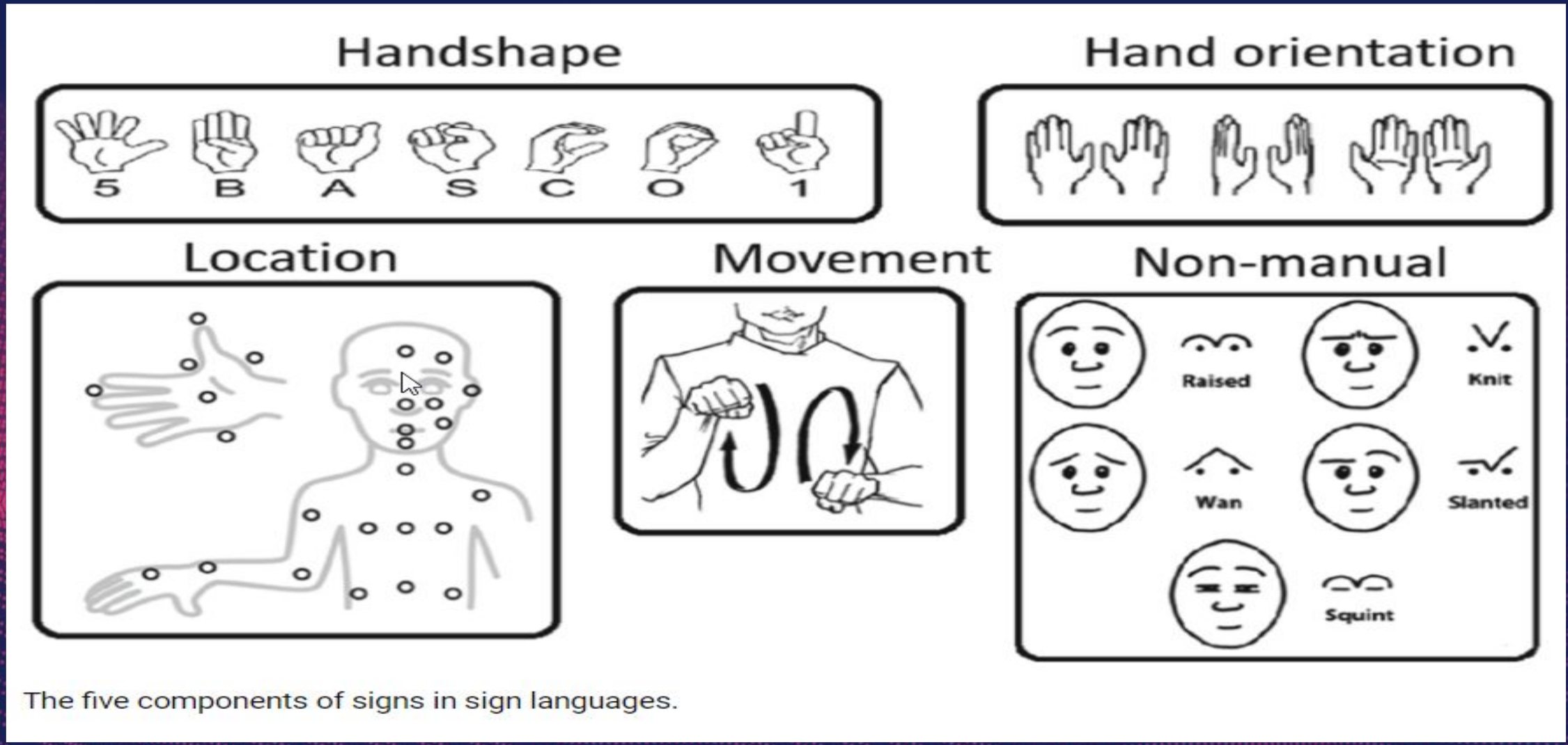
The sign language is used by those who are hearing impaired as a medium of communication. Sign Language is composed of various hand gestures, movements, orientation and facial expressions.

Mission Statement

Building Inclusive Communities for All

Our mission is to break the communication barriers for hearing impaired communities by creating an effective and accessible sign language interpretation solution.

Components of Sign Language : Multimodal



The five components of signs in sign languages.

ASL Interpretation: high demand but limited access

Target population

- **Approximately more than a half-million people throughout the US use ASL to communicate as their native language.**

Limited Employment

- **3.8% are unemployed**
- **42.9% are not in labor force.**

Limited Access

- **American sign language interpreters charge and average of $200 per hour.**
- **Travel costs are additional costs paid by requestors.**
- **In most cases, it is necessary to book an interpreter for a minimum period of 2-hours.**

10,253 certified ASL interpreters in the US and Canada (Registry of interpreters for hearing impaired)

ASL Interpretation – Important in many industries and required by law

01

Main Sectors

- Healthcare
- Education
- Employment
- Social Services
- Legal
- Entertainment
- Government

02

Regulations

- Early Hearing Detection and Intervention
- Individuals with Disabilities Education Act
- No Child Left Behind Act
- Rehabilitation Act of 1973
- Americans with Disabilities Act
- Fair Housing Act
- Television Decoder Circuitry Act
- Air Carrier Access Act
- Communications Act

We are building a machine-learning enabled solution to increase access of ASL interpretation

Solution

Continuous sign language recognition, interpretation and translation in real-time.

Beneficiary

The hearing and speaking impaired communities.

Social Impact

increased access to information and communication, fostering inclusive society.

Demonstration with a Scenario

Scenario	Conversation	Signs
Classroom Setting	Hello teacher, love the class.	• hello • teacher • love • class

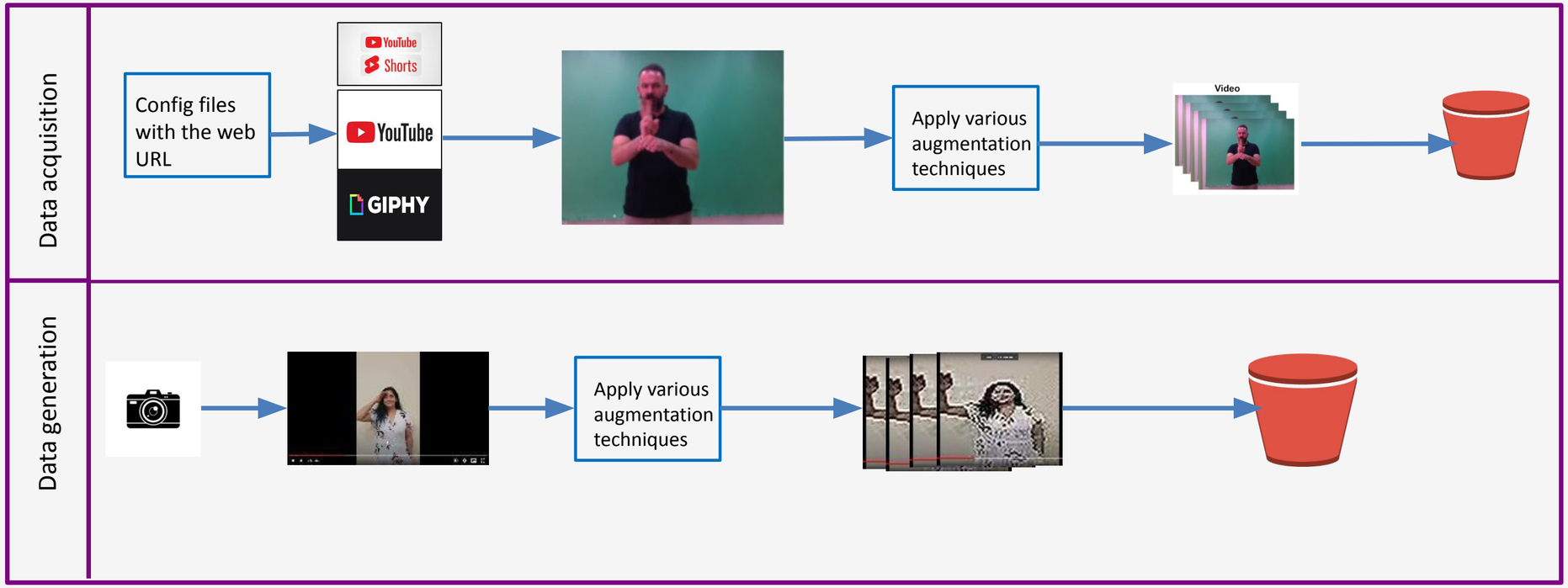
Technical Solution Overview

Capabilities we have built:

- ASL video data acquisition and generation
- Data cleaning, processing, and transformation
- Data pipeline for multi-class classification
-
- Multi-pronged approach to modeling
-
- Model Evaluation
- Real time Model Inference on a full ASL video
- End to end deployment with interpretation and deployment

.

ASL Video Data Acquisition and Generation Pipeline



ASL video data challenges: data Acquisition

Data Acquisition

- lack of standards in ASL it is extremely difficult to get quality datasets.
- In addition, the resources it takes do the language annotation and due to scarcity of annotators, there are very limited public datasets available.
- The only solution was to perform web scraped data generation for a subset of isolated signs.
- Created scenario table with real life conversations so that we can narrow our data gathering effort

ASL video data generation

Data Generation

- Recorded individual signs for the scenario table to experiment and compare the model inferencing results.
- **We created a set protocol to record our videos.**
  - The cue was to start recording after 30 seconds of the start of the video using a timer.
  - Perform a sign and wait 30 seconds on a timer to have consistency.
- Though we are able to record a number of isolated signs, as we are not professional signers, it was extremely challenging to get the syntax right.

Model Experimentations

Mediapipe + LSTM

Slow and resource intensive

ConvLSTM

Did not generalize well

Conv2plus1D

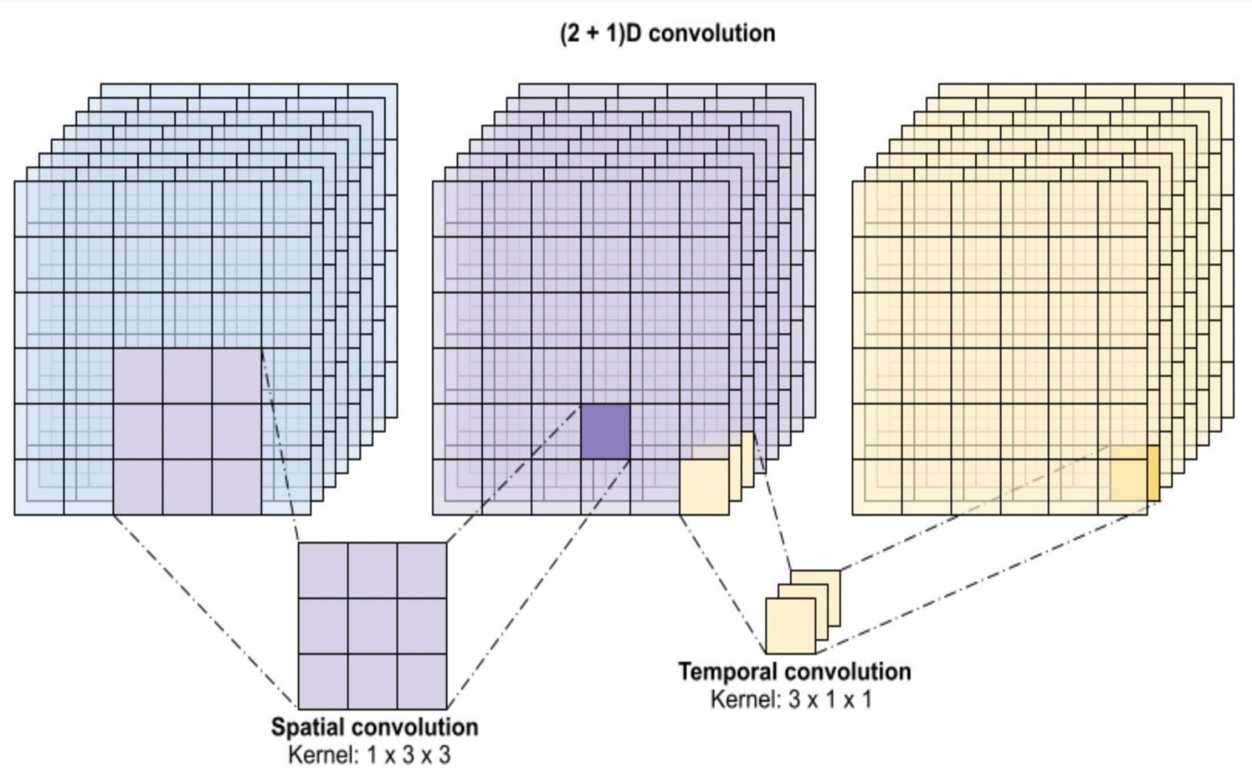
Performed better than previous 2

Winning Model

MoViNet

Transfer learning model, provided the best results

Model Consideration (2 + 1)D Convolution



- Considers both spatial and temporal factors in the video
- Reduced number of parameters
- The input video data is in avi format.

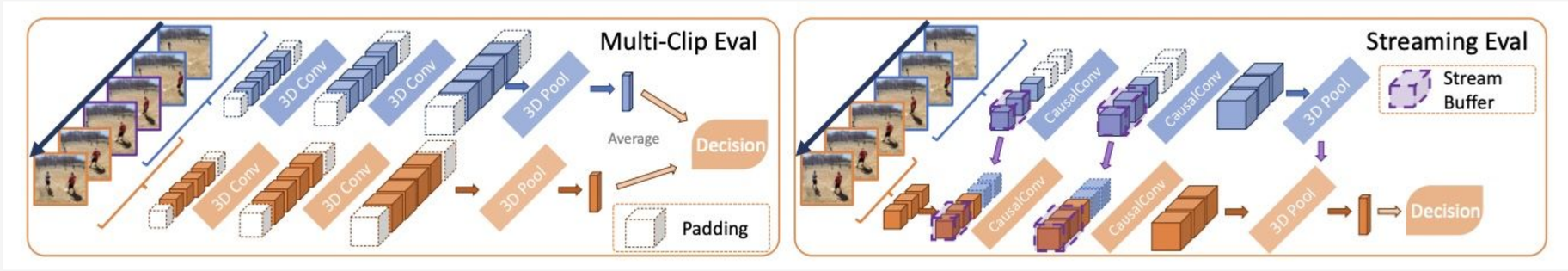
(2 + 1)D Convolution Model Performance

- We have tested Conv 2plus1d model with 5 and 8 classes.
- We trained our model separately with acquired, generated and hybrid data.
- Better performance with acquired over generated data.

(2 + 1)D Convolution Model Performance

Model	Experiment No	Number of Signs	Data	Number of Epochs	Learning Rate	Training Accuracy	Validation Accuracy	Test Accuracy
conv2plus1d	1	5	Aquiried	50	0.0001	64%	65%	58%
	2	5	Aquiried	75	0.0001	77%	67%	52%
	3	5	Aquiried	100	0.0001	83%	73%	70%
	4	5	Aquiried	50	0.001	54%	55%	50%
	5	5	Aquiried	75	0.001	57%	53%	40%
	6	5	Aquiried	100	0.001	73%	60%	45%
	7	5	Generated	50	0.0001	51%	38%	52%
	8	5	Generated	75	0.0001	53%	33%	48%
	9	5	Generated	100	0.0001	59%	37%	50%
	10	5	Generated	50	0.001	39%	23%	28%
	11	5	Generated	75	0.001	41%	35%	38%
	12	5	Generated	100	0.001	53%	37%	40%
	13	8	Aquiried+Generated	50	0.0001	40%	38%	32%
	14	8	Aquiried+Generated	75	0.0001	50%	41%	40%
	15	8	Aquiried+Generated	100	0.0001	67%	55%	52%
	16	8	Aquiried+Generated	50	0.001	31%	28%	25%
	17	8	Aquiried+Generated	75	0.001	32%	27%	25%
	18	8	Aquiried+Generated	100	0.001	44%	33%	30%
	19	5	Aquiried+Generated	50	0.0001	60%	51%	46%
	20	5	Aquiried+Generated	75	0.0001	65%	53%	57%
	21	5	Aquiried+Generated	100	0.0001	68%	53%	52%
	22	5	Aquiried+Generated	50	0.001	38%	36%	32%
	23	5	Aquiried+Generated	75	0.001	49%	31%	29%
	24	5	Aquiried+Generated	100	0.001	45%	42%	40%

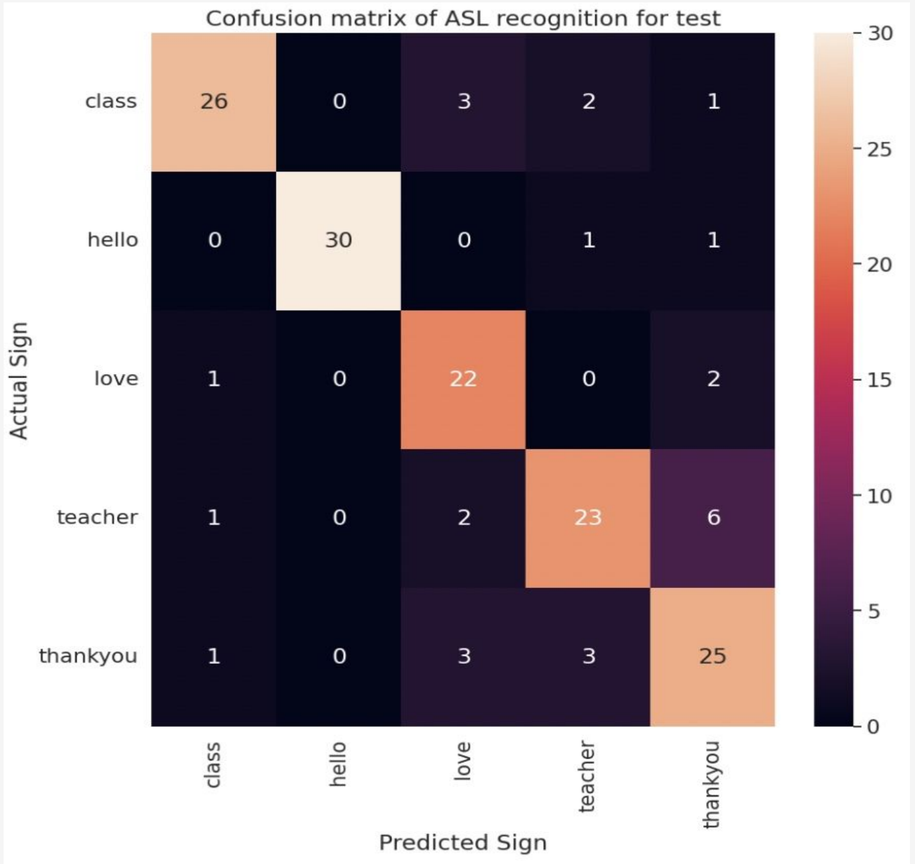
Model Consideration MoViNet Base and Streaming



- Adopted after Mobile Video Networks for Efficient Video Recognition
- Supports frame by frame inference
- MoViNets are more accurate than 2D networks and more efficient than 3D networks.
- MoViNets are a family of memory and computation efficient 3D CNNs algorithms

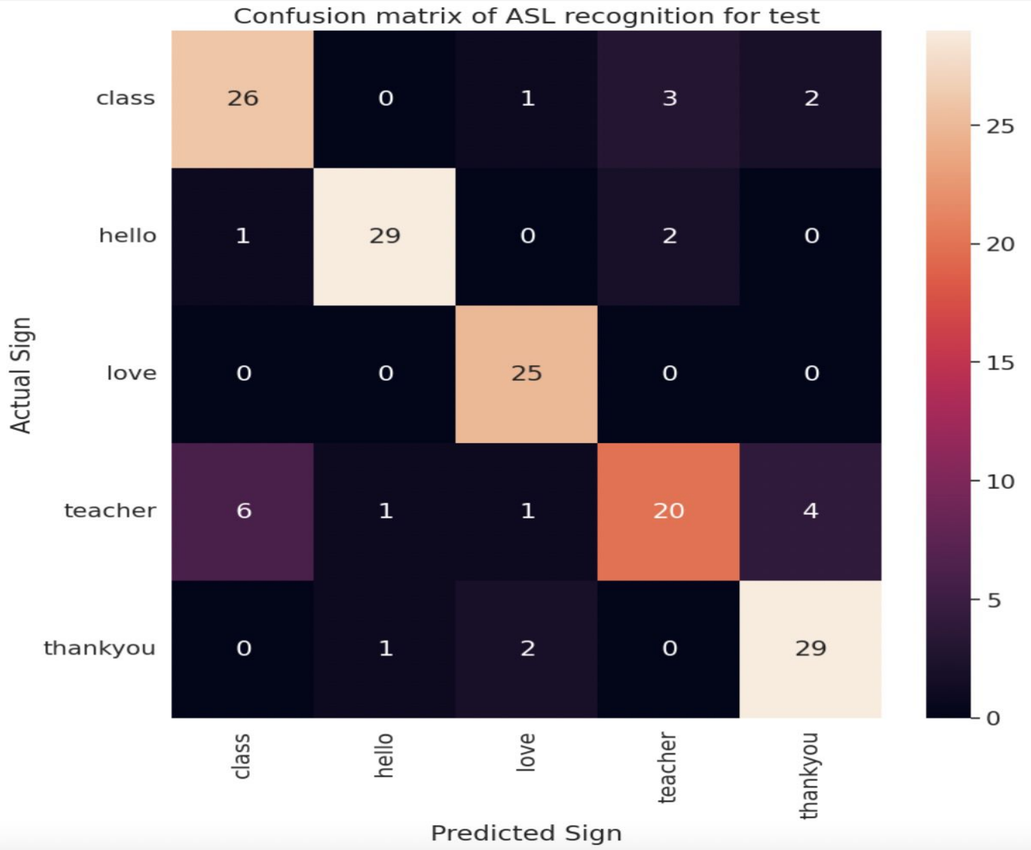
MoViNet Base Performance: a0 Model

	Precision	Recall	F1	Support
Class	0.9	0.81	0.85	32
Hello	1	0.94	0.97	32
Love	0.73	0.88	0.8	25
Teacher	0.79	0.72	0.75	32
Thank you	0.71	0.78	0.75	32
Accuracy			0.82	
Macro Avg	0.81	0.83	0.82	153
Weighted Avg	0.81	0.82	0.83	153

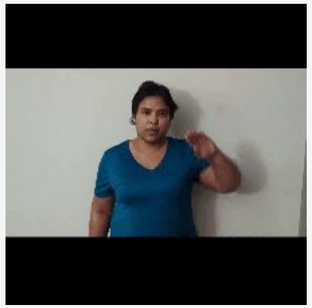


MoViNet Streaming Performance:a1 Model

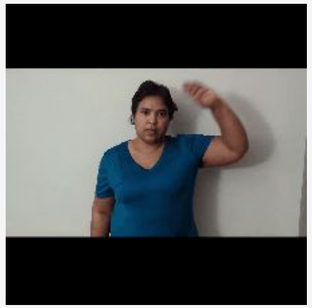
	Precision	Recall	F1	Support
Class	0.79	0.81	0.8	32
Hello	0.94	0.91	0.92	32
Love	0.86	1	0.93	25
Teacher	0.8	0.62	0.7	32
Thank you	0.83	0.91	0.87	32
Accuracy			0.84	
Macro Avg	0.84	0.85	0.84	153
Weighted Avg	0.84	0.84	0.84	153



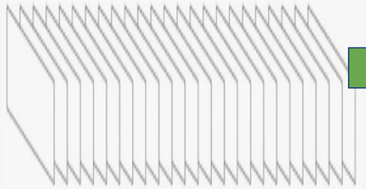
MoViNet Inference Pipeline



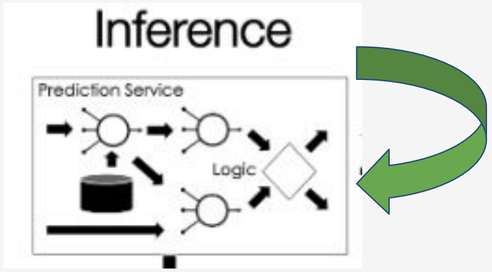
Original Input Video



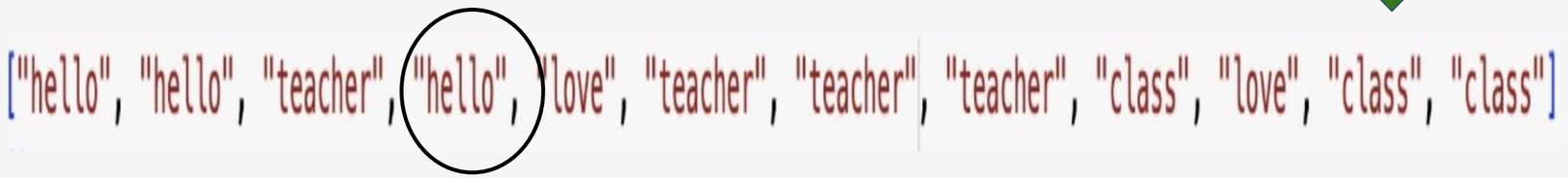
Reduced Sized Video



Video Segments



Inference Model



Grammar Error Correction (GEC)

- The inference model transcribes the English gloss, which is not yet a grammatically correct sentence.
- So we used **finetuned-llama-2-70b** grammar correction module which uses the inference model output gloss and translates into a grammatically correct English sentence.

```
["hello", "hello", "teacher", "hello", "love", "teacher", "teacher", "teacher", "class", "love", "class", "class"]
```

Inference Model Output

```
{'correction': 'Hello Teacher, I love class!'}
```

GEC Output in American English

Multilingual Translation using facebook/m2m100_418M

```
{'correction': 'Hello Teacher, I love class!'}
```

English Text

```
[{'generated_text': 'Bonjour professeur, j'adore les cours !'}]
```

French Text

```
[{'generated_text': 'हाय शिक्षक, मुझे कक्षा पसंद है!'}]
```

Hindi Text

```
[{'generated_text': 'Hola profesora, me encanta la clase!'}]
```

Spanish Text

Analysis

In our experiments we have few major learnings:-

1. Data is the key. Quality of the data dictates the model performance.
2. Variability of the signers can change the model performance a lot.
3. Clear background with no background objects provide better model performance.
4. Lower learning rate along with lot of training time produces better results.
5. This is a resource hungry process. We need lot of data to get good results.
6. Inference model with trained weights do not work same way as simple model save or model save weights. Rather the model weights need to be saved in tflite format to be used later.

Future Work & Roadmap

1. We want to extend our work with new signs and test the model performance.
2. We would also like to evaluate model performance in a continuous setting using both signs and finger spelling.
3. We want to extend the model architecture to adapt to other sign languages like Indian Sign Language.
4. We want to test the performance in the mobile application as well.

Mission Statement

Building Inclusive Communities for All

Our mission is to break the communication barriers for hearing impaired communities by creating an effective and accessible sign language interpretation solution.

ASL interpretation has high demand, but access is limited. It is used in various sectors and are required by federal laws.

Our machine-learning empowered solution aims to increase access of ASL interpretation and breaks the communication barrier for the hearing and speaking impaired communities.

Anknowledgements

1. We would like to thank our instructors Joyce, Kira, Mark Butler, Alex D for all their support, guidance encouragements and references.
2. We would like to express our thanks to our teaching assistants Prabhu, Dannie and Jordan for their help in various project phases.
3. We want to thank subject matter expert Jenny Buechner and Haya Naser for their time and guidance.
4. We would like to express our thanks to our classmate Olivia Pratt from DATASCI (Data Science) 231: Behind The Data: Humans And Values for her in depth analysis about model fairness.
5. Finally, we are grateful to our friends and families for their unwavering support, guidance, encouragement to reach to the finish line.

Referrences

1. https://cdhh.ri.gov/information-referral/american-sign-language.php#:~:text=ASL%2C%20short%20for%20American%20Sign,communicate%20as%20their%20native%20language.
2. https://www.handtalk.me/en/blog/universal-sign-languages/#:~:text=It%20may%20come%20as%20a,and%20parts%20of%20Southeast%20Asia.
3. https://www.123rf.com/photo_88216121_isolated-deaf-icon-symbol-on-clean-background-vector-mute-element-in-trendy-style.html
4. https://www.vecteezy.com/vector-art/9684134-vector-sign-of-the-percentage-symbol-is-isolated-on-a-white-background-percentage-icon-color-editable
5. https://www.statista.com/statistics/1095081/employment-unemployment-labor-force-rates-deaf-and-hearing-us/
6. https://www.istockphoto.com/vector/vector-image-of-a-flat-isolated-icon-dollar-sign-currency-exchange-dollar-united-gm1151557689-312128949
7. https://languagers.com/video-remote-interpretation-services-how-much-does-vri-cost/
8. https://www.visualpharm.com/free-icons/percentage-595b40b85ba036ed117dc34b
9. https://www.nad.org/resources/american-sign-language/interpreting-american-sign-language/#:~:text=The%20demand%20for%20qualified%20interpreters,remote%20interpreting%20(VRI)%20services.
10. https://www.nad.org/resources/civil-rights-laws/early-hearing-detection-and-intervention/
11. https://www.etsy.com/listing/1511120611/camera-svg-vintage-camera-silhouette?gpla=1&gao=1&&utm_source=google&utm_medium=cpc&utm_campaign=shopping_us_e-craft_supplies_and_tools-canvas_and_surfaces-stencils_templates_and_transfers-clip_art&utm_custom1=_k_Cj0KCQiAyeWrBhDDARIsAGP1mWRISMPMWfQYvEeJ65fHKRMG-MaZUGP9dXuwgOHXUTamnjtA0-3MZOAaAvm7EALw_wcB_k_&utm_content=go_12564966396_119881495815_507186268783_pla-295943621186_c__1511120611_471013805&utm_custom2=12564966396&gad_source=1&gclid=Cj0KCQiAyeWrBhDDARIsAGP1mWRISMPMWfQYvEeJ65fHKRMG-MaZUGP9dXuwgOHXUTamnjtA0-3MZOAaAvm7EALw_wcB
12. https://www.creativefabrica.com/product/video-26/
13. https://thenounproject.com/browse/icons/term/data-repository/
14. https://arxiv.org/abs/1711.11248v3
15. https://arxiv.org/pdf/2103.11511.pdf