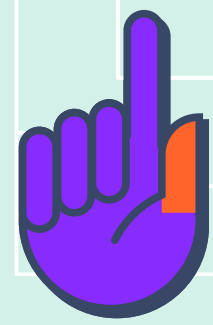




CSE 4554

Word-level Sign Language Recognition Through Deep Learning



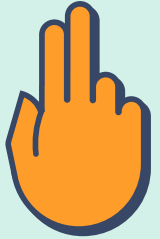
Exploring Sign Language Recognition Through LSTM and Cosine Similarity Modeling

[Github Link](#)





Team Members



**Nafisa
Maliyat**

ID: 200042133



**Shanta
Maria**

ID: 200042172



Contents

01 Problem
Statement

02 Solution

03 Model
Architecture

04 Dataset

05 Feature
Engineering

06 Training





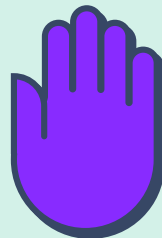
Contents

07 Result

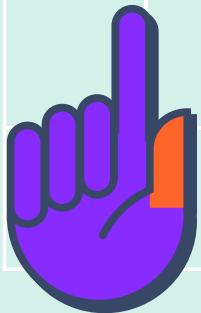
09 Challenges

08 Analysis

10 Future Work



01 Problem Statement

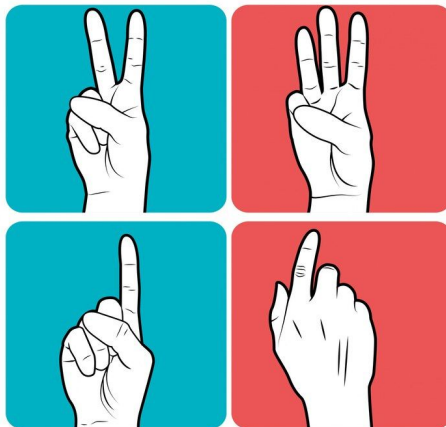




**Human
Translator
Expensive**



**Extensive
Vocabulary &
Sign Variations**



**Communication
Barrier**



02 Solution



**Word Level
Translation**

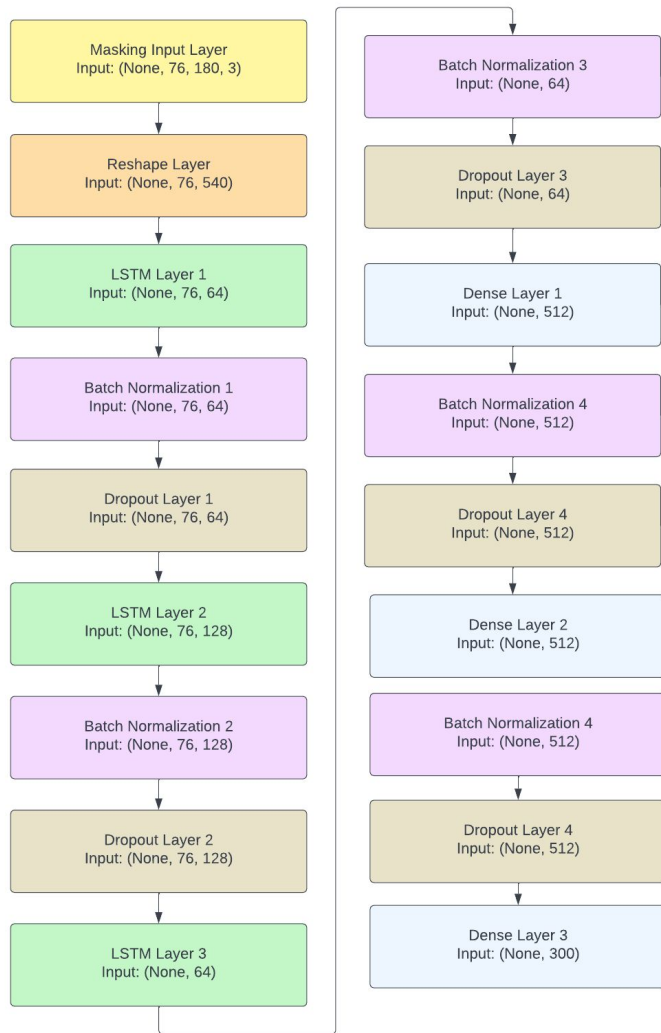
**Deep Learning,
MediaPipe,
Cosine Similarity**

**Focus on
Interpretation of
Fine-Grained
Gestures**

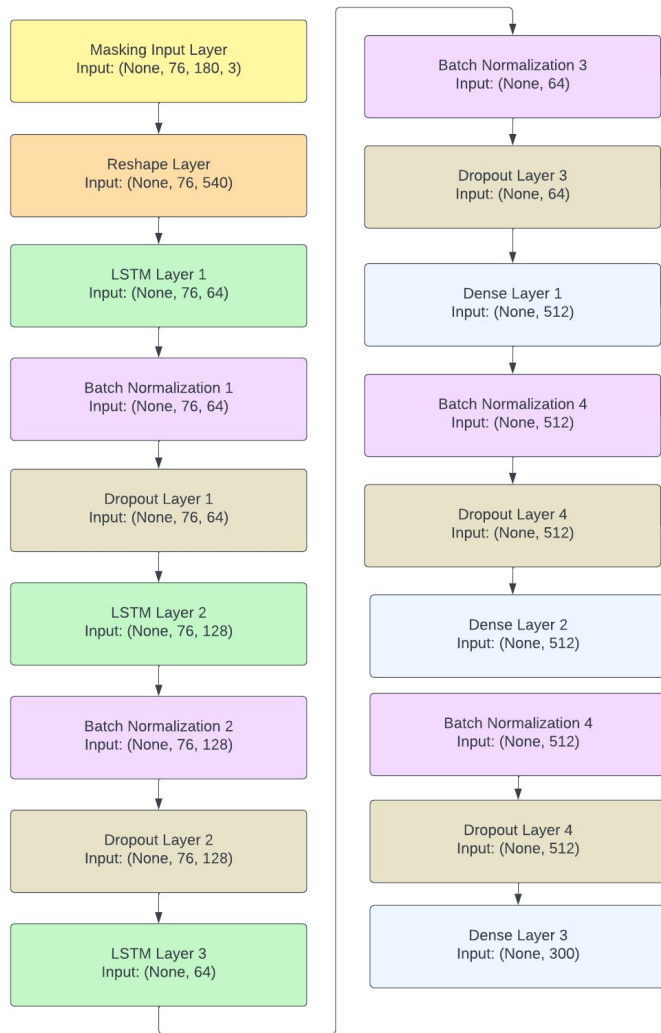


03 Model Architecture

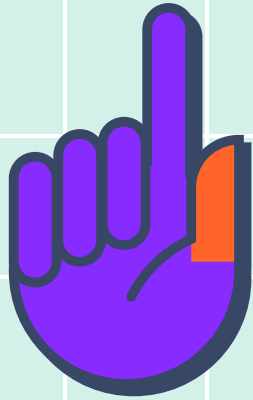




- **Input Layer:**
 - Masking Layer = ignores padded values
- **Reshape Layer:**
 - Convert to 3D format
- **LSTM Layers:**
 - 3 Layers
 - ReLu Activation
 - First 2 Layers have return sequences

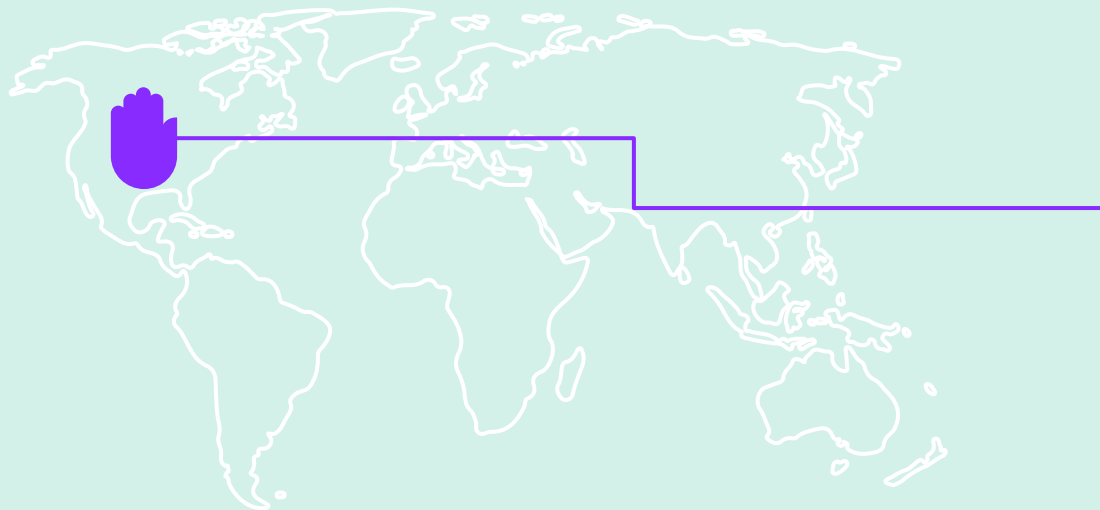


- **Dense Layers:**
 - 2 Layers
 - ReLu Activation
- **Output Layer:**
 - Dense Layer
 - Linear activation
- **Techniques:**
 - Batch Normalization
 - Dropout (rate=0.5) = Regularization
 - Applied after each hidden layer



04 Dataset

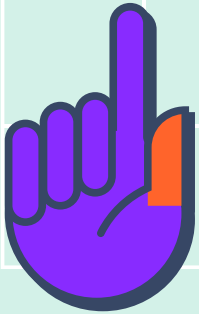
Dataset



WLASL

- 21,083 videos corresponding to 2,000 common ASL words
- Metadata provided

Feature Engineering 05



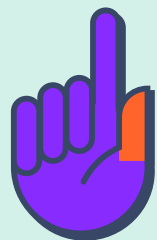


Landmark Extraction

- **Hands:**
 - 42 (21 on each hand)
- **Pose:**
 - 6 landmarks
 - For the upper body
 - Excluding the face
- **Face:**
 - 132 out of the 478
 - Focusing on the lips, eyes, eyebrows, and the outline of the face

Total Landmarks: 180
Landmarks Centered





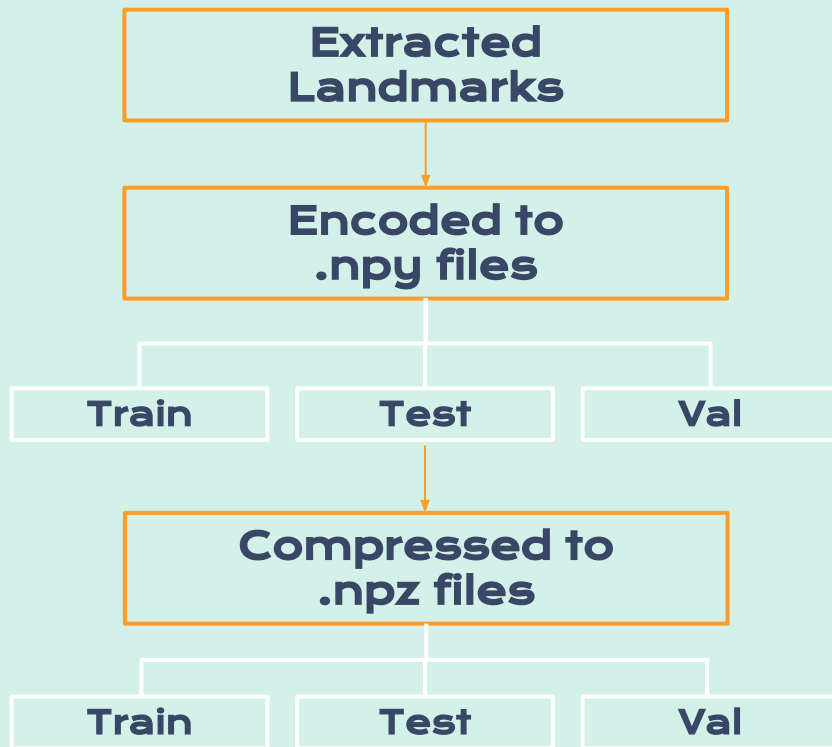
Landmark Extraction

Body	Hands (Left and Right)	Face
Left Shoulder Right Shoulder Left Elbow Right Elbow Left Wrist Right Wrist	Wrist Thumb CMC Thumb MCP Thumb IP Thumb Tip Index Finger MCP Index Finger PIP Index Finger DIP Index Finger Tip Middle Finger MCP Middle Finger PIP Middle Finger DIP Middle Finger Tip Ring Finger MCP Ring Finger PIP Ring Finger DIP Ring Finger Tip Pinky MCP Pinky PIP Pinky DIP Pinky Tip	Face Outline Lips Eyes Eyebrows





Preprocessing



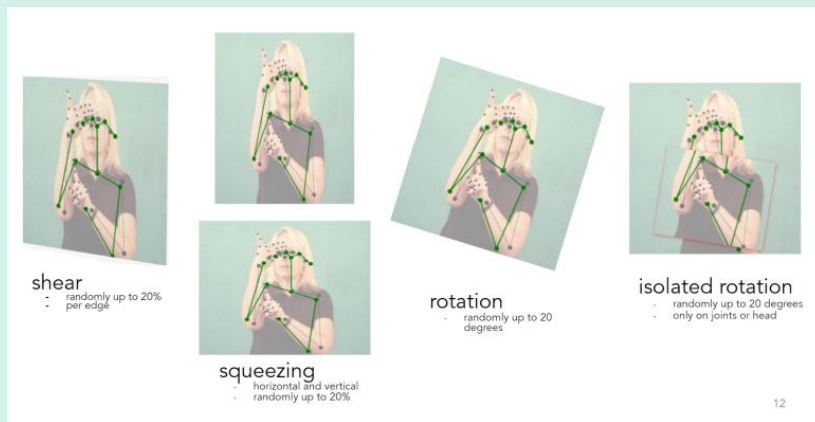


Augmentation

Rotation

Zoom

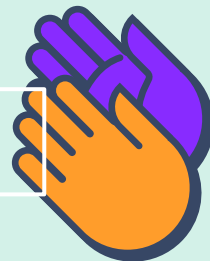
Shift



Mask

HFlip

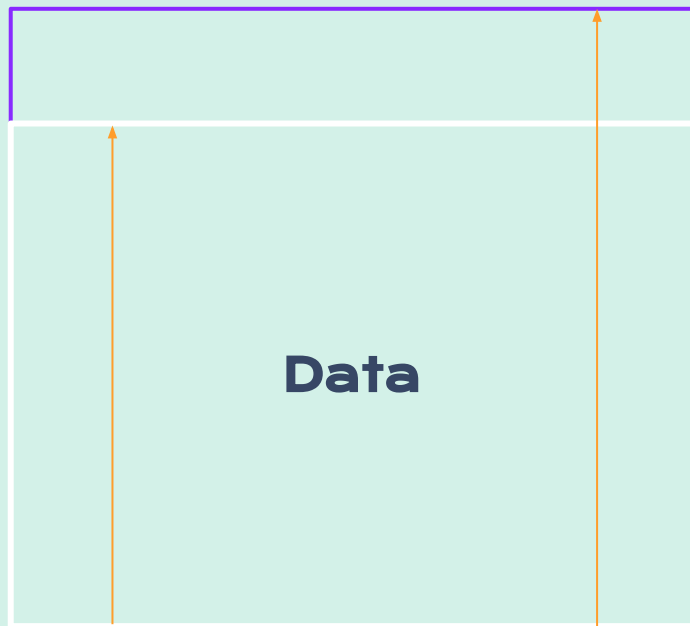
Speedup





Padding

X=74
(example)

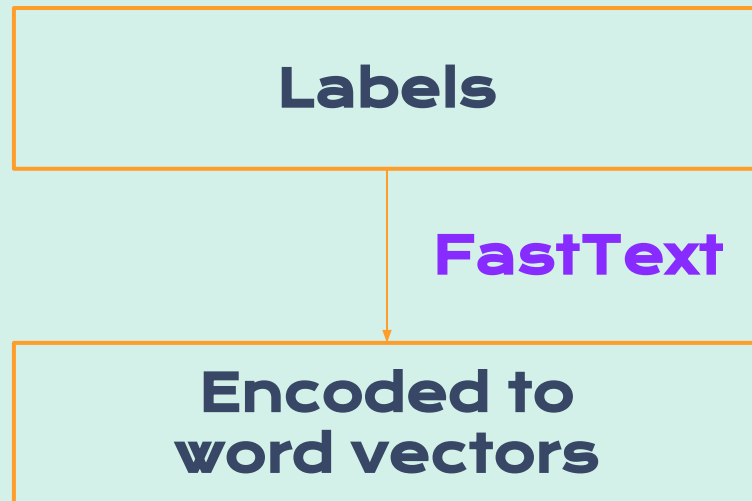


X=76
(expected)

**Padding
value = -80**
for $(76-74=)$
2 frames
each with
180
landmarks



Label Encoding



06 Training





**Batch
Size = 128**



Epochs = 100



Validation

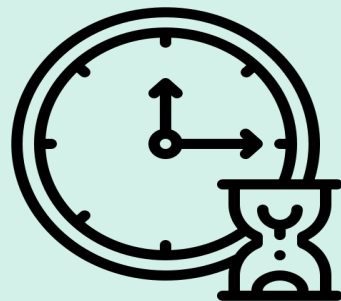




Early Stopping

```
early_stopping = tf.keras.callbacks.EarlyStopping(  
    monitor='val_accuracy',  
    patience=10,  
    restore_best_weights=True  
)
```

Patience = 10





Model Checkpoint



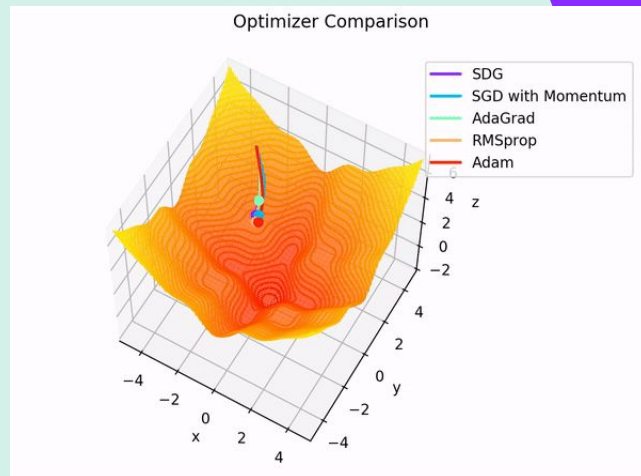
```
model_checkpoint = tf.keras.callbacks.ModelCheckpoint(  
    filepath=os.path.join(checkpoint_filepath, 'model_{epoch:02d}.h5'),  
    save_weights_only=True,  
    save_best_only=True,  
    monitor='val_accuracy',  
    mode='max',  
)
```




Optimizer

Adam Optimizer ($\alpha = 0.05$, 0.9 decay every 1000 steps)

```
lr_scheduler = tf.keras.optimizers.schedules.ExponentialDecay(  
    initial_learning_rate=0.05,  
    decay_steps=1000, # decrease learning rate at an exponential rate  
    decay_rate=0.9  
)
```

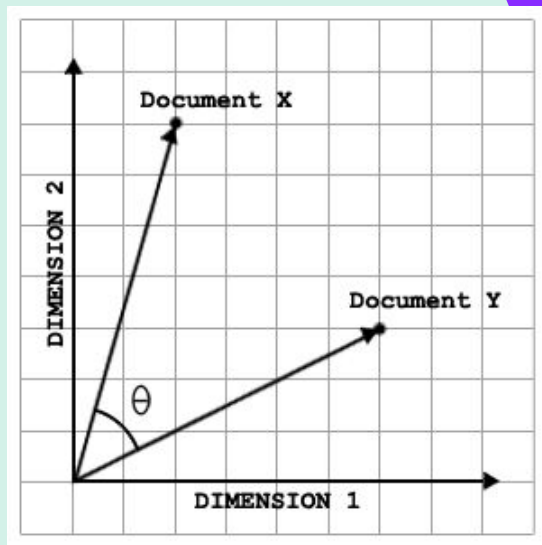




Loss

Cosine Similarity Loss

```
model.compile(  
    loss=tf.keras.losses.cosine_similarity,  
    optimizer=tf.keras.optimizers.Adam(learning_rate=lr_scheduler),  
    metrics=['accuracy']  
)
```





07 Result



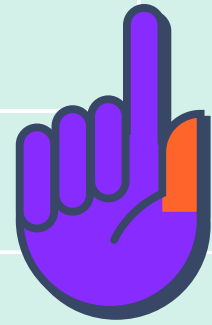


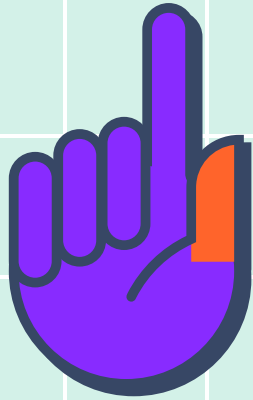
16.68%

Test Accuracy on

2000

labels i.e. the entire dataset





08

Analysis



Prediction

Accuracy

Total Accuracy VS Total
Validation Accuracy



Loss

Total Loss VS Total
Validation Loss



Confusion Matrix

For first 50 samples



ROC

For each unique label
with AUC





Subset Sample Size	Test Accuracy (%)
50	22.378
200	21.154
1000	17.964
2000	16.678

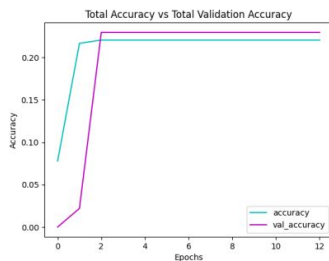


Fig. 5: For 50 samples

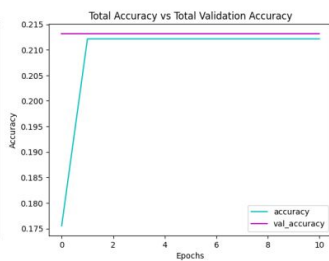


Fig. 6: For 200 samples

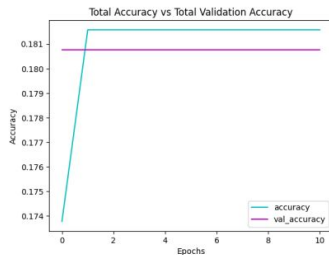


Fig. 7: For 1000 samples

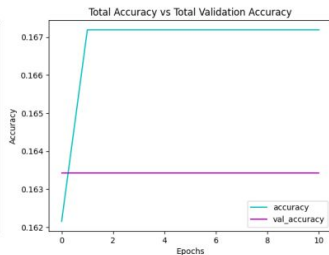


Fig. 8: For 2000 samples

Fig. 9: Total Accuracy vs Total Validation Accuracy

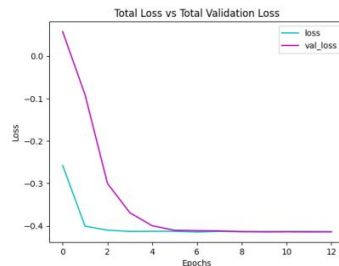


Fig. 10: For 50 samples

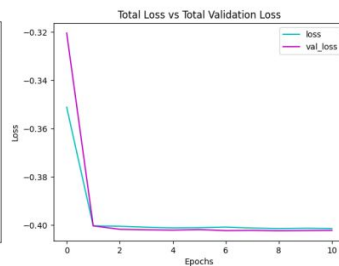


Fig. 11: For 200 samples

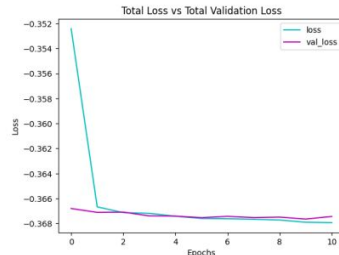


Fig. 12: For 1000 samples

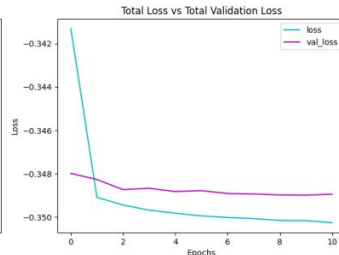
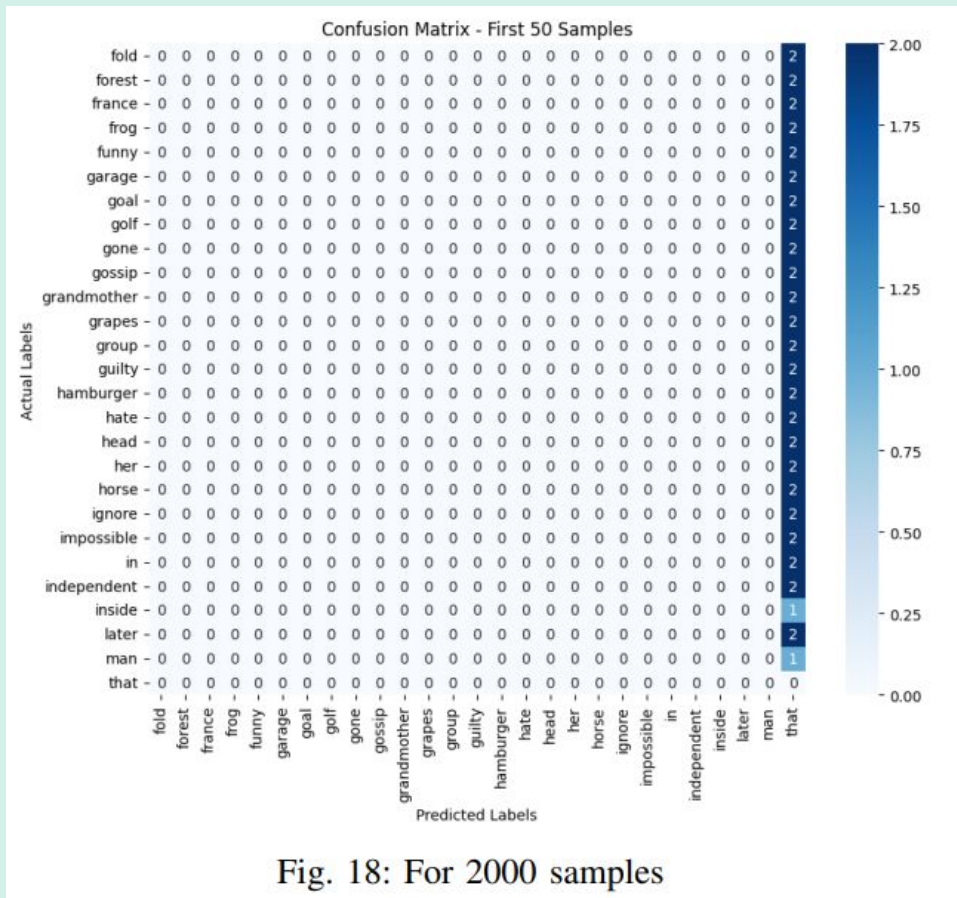


Fig. 13: For 2000 samples

Fig. 14: Total Loss vs Total Validation Loss





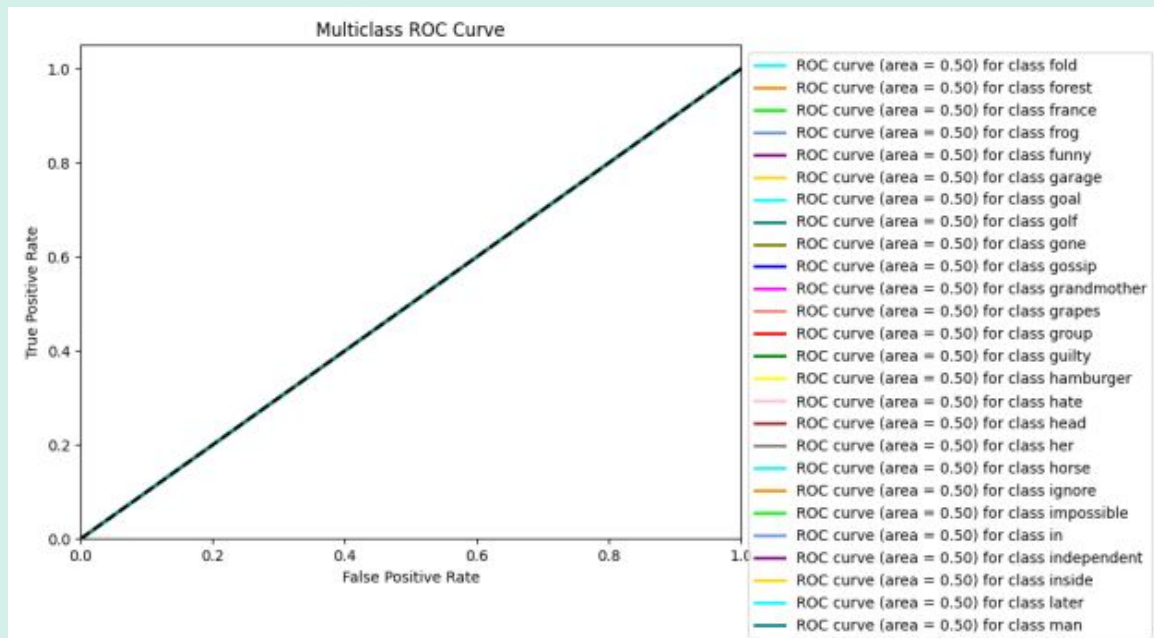
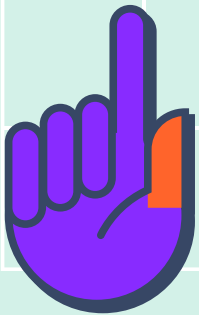


Fig. 23: For 2000 samples



Challenges 09

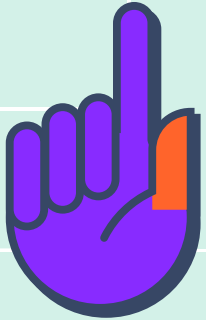


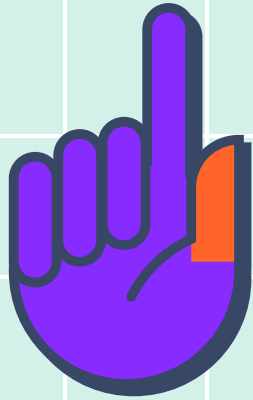
Small epoch

**Computational resource
constraint**

Time constraint

Padding & Augmentation





10 Future Work



**Larger
Epoch**

**More
Augmentation**

**Real-Time
Translation**

**Diverse &
Larger
Dataset**

**Sentence
Level
Translation**

**Thank
You For
Your
Patience**

