# DSCI 6011-2: Deep Learning Project

# Comparison of Sign Language Detection Architecture

- Bikash  Adhikari
- Devanshi Tandel
- Ijeoma E. Chukwuma
- Medha kanu Baniya

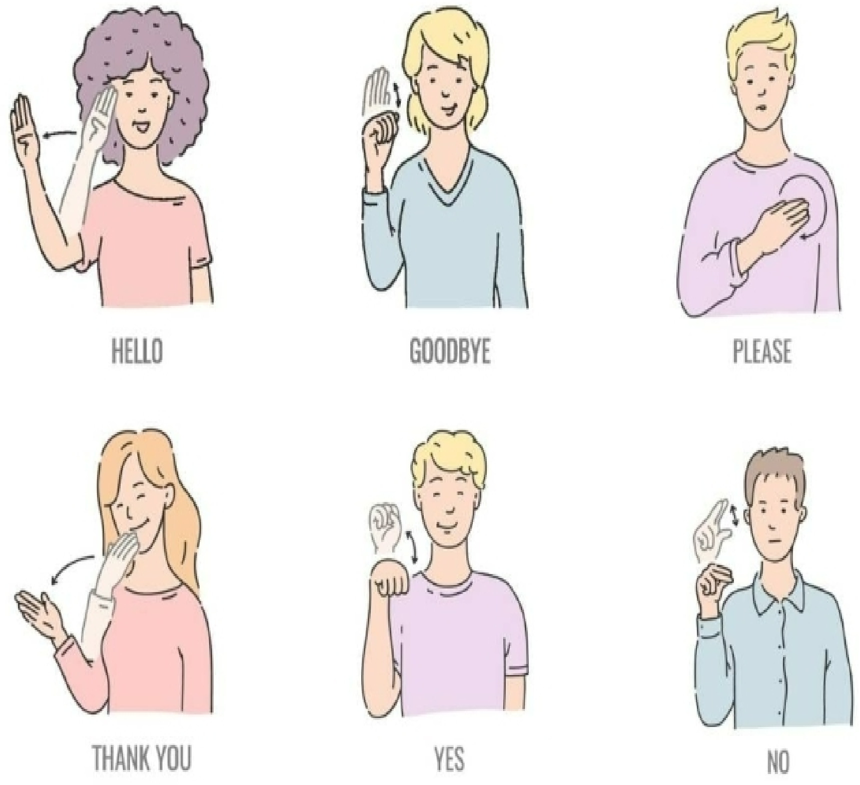
# Statement of Project Objectives

Compare the predictive accuracy of YOLOv5 and Vision Transformer model in recognizing gestures/signs with multiple meanings and translating these gestures into text or speech, ensuring real-time efficiency for practical application.

# Statement of Value

- Sign language is a visual-manual form of communication, using hand gestures, facial expressions, and body language. It's a rich linguistic system with its own grammar and syntax, distinct from spoken languages. Each gesture conveys words, concepts, or sentences, enabling nuanced communication beyond spoken words.
- This acknowledgment highlights its importance not just for the Deaf and hard-of-hearing community but also as a vital element of linguistic diversity and cultural richness.

https://accessibe.com/glossary/sign-language



# Review of the State of the Art and Relevant Works

Research in sign language detection has primarily focused on developing models that accurately recognize individual signs using various deep learning architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). These studies have achieved significant success in sign recognition accuracy. However, a noted gap in the literature is the challenge of interpreting signs with multiple meanings based on context. Current methodologies cannot often incorporate contextual clues that are crucial for understanding the intended meaning of a sign, limiting the applicability of these technologies in real-world communication scenarios.

IM. Al-Qurishi, T. Khalid and R. Souissi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," in IEEE Access, vol. 9, pp. 126917-126951, 2021, doi: 10.1109/ACCESS.2021.3110912.

# Review of the state of the art and relevant works

Future research can enhance model accuracy through diverse datasets, camera orientation adjustments, and wearable device incorporation. The current models focus on isolated signs, which can be utilized for interpreting continuous sign language, leading to syntax generation in ISL. Vision transformers can provide more accurate results than feedback-based learning models.

Kothadiya D, Bhatt C, Sapariya K, Patel K, Gil-González A-B, Corchado JM. Deepsign: Sign Language Detection and Recognition Using Deep Learning. *Electronics*. 2022; 11(11):1780. https://doi.org/10.3390/electronics11111780

# Approach

We will take the input as a photo/ video signal, perform the comparison between the proposed models, and show the result from both model side by side as a method of evaluation.

- Our proposed dataset is -> https://www.kaggle.com/code/hengck23/lb-0-67-one-pytorch-transformer-solution/input
- Algorithm and Model: Comparison between Yolov5 and Vision Transformer
- Tools and Techniques: AWS for deployment

# Deliverables

GitHub Repository

Web Application

Report
(Comparison of
Different Models )

Presentation File

# Evaluation Methodology

- Output of proposed model side-by-side
- For both model comparisons :
  - ➢ F-1 Score: We will use it to evaluate the models' overall accuracy.
  - ➢ Speed: Time needed for the models to make a prediction.
  - ➢ Robustness:  We will test and compare in different scenarios for the model performance consistency.

T
H
A
N
K
Y
O
U