# Sign Language Prediction Software

**Abstract:**

This project focuses on the development of a caption generation system tailored for top hand sign language interpretation, integrated into a bipedal humanoid robot. Utilizing deep learning techniques, the humanoid interprets visual hand gestures and generates meaningful textual descriptions. The system merges computer vision with natural language processing (NLP) to enhance real-time communication, improving accessibility and interaction between humans and machines, especially for the hearing and speech impaired.

**Project Objectives:**

* Develop a vision-based caption generator for interpreting top hand sign language in a humanoid robot.
* Enable the robot to describe recognized hand gestures using deep learning.
* Enhance human-robot communication by translating gestures into contextual language.
* Optimize the model for real-time performance on embedded platforms.

**Methodology:**

The system operates in two main stages: visual gesture recognition and caption generation.

1. **Gesture Recognition (Feature Extraction):**  
   A pre-trained Convolutional Neural Network (CNN), such as InceptionV3 or MobileNet, captures and processes hand gestures from the robot's camera input to extract feature vectors.
2. **Caption Generation:**  
   The extracted features are passed to a Long Short-Term Memory (LSTM) network trained on a hand sign-caption dataset to produce corresponding textual output.
3. **Speech Integration:**  
   Captions are converted to audio using a Text-to-Speech (TTS) engine, enabling spoken feedback for enhanced interaction.

**Key Findings:**

* The model accurately interprets and captions top hand signs in real time.
* Lightweight CNN architectures like MobileNet allow for efficient execution on embedded systems.
* Transformer-based models (e.g., Vision Transformers or CLIP) offer potential accuracy gains over traditional LSTM models.
* Integration with TTS significantly enhances user experience, especially in accessibility contexts.

**Step-wise Solution Approach:**

1. **Data Collection and Preprocessing:**
   * Compile a labeled dataset of top hand sign images with corresponding textual meanings.
   * Preprocess images (resizing, normalization) and tokenize captions for model training.
2. **Feature Extraction:**
   * Use CNN models (e.g., InceptionV3, MobileNet) to extract gesture-specific features from image inputs.
3. **Caption Generation Model:**
   * Train an LSTM or Transformer-based model to generate meaningful captions from gesture features.
4. **Model Optimization:**
   * Tune hyperparameters to achieve an optimal trade-off between accuracy and inference speed.
   * Adapt the model for deployment on hardware such as Raspberry Pi or Jetson Nano.
5. **Evaluation:**
   * Evaluate caption accuracy using BLEU, METEOR, and CIDEr metrics.
   * Perform real-world testing with users to assess reliability and interaction quality.
6. **Deployment in Humanoid:**
   * Integrate the gesture recognition and captioning model into the humanoid's vision and control system.
   * Enable audio output via a TTS engine to vocalize the interpreted hand signs.
   * Test for robustness and real-time performance in diverse environments.
7. **Team Members:**

Anand gupta – E22CSEU1352

Bal Munkund - E22CSEU1352