

# “Mudra” (Text to Speech Conversion with Hand Gestures)

*Prof. Poonam Lad, Rohit Pralhad Katkar, Vishal Balasaheb Bochare, Srikant Santosh Jampal,  
Hrushikesh Dhanaji Tambe  
Department of Information & Technology.  
Pillai HOC College of Engineering & Technology, Rasayani.*

## I. ABSTRACT

One of the non-verbal communication techniques utilized in sign language is the hand gesture. Deaf persons who struggle with speech or hearing most frequently use it to converse with each other and hearing people. Several makers around the world have created various sign language systems, but they are neither adaptable nor cost-effective for end users. The majority of programs only provide text-based results. So, we are going to offer a system prototype in an application that solves the aforementioned issue. Hand gestures can be translated into text and then spoken by the system. Several algorithms will be used by the system, including CNN for classifying hand movements and SVM for detection of hand movement & TTS for conversion of Text To Speech.

## II. INTRODUCTION

The increase in general acceptance and funding of international projects emphasizes the need for sign language. In this day and age of deaf technology, it's important to want a computerized solution. However, researchers have been working on the issue for a while, and the outcomes are encouraging. While there are intriguing voice recognition technologies available, there are no commercial character recognition solutions available at this time. The objective is to create a Human Computer Interface (HCI) that can comprehend human speech. Teaching computers to recognize speech, facial expressions, and human gestures is one step toward this objective. Information is conveyed nonverbally through gesturing. Gesture recognition is an aspect of human-computer interaction that demonstrates an academic thesis. It is essential for popularizing the

idea of human connection, an open dialogue that must include user-machine correlation. Motion investigation is a discipline that can perceive signals, for example, hand, arm, head and, surprisingly, primary developments that normally include a specific stance or potentially development. A person can send more information in less time by using hand gestures. Computer vision concepts have been applied to the real-time processing of gesture outputs in a number of different ways.

## III. REQUIREMENTS

### 1. Software Requirements

- Operating System: Windows 10 or Later
- Android Studio With JAVA
- Programming language : Python for machine learning.

### 2. Hardware Requirements

- Processor: Intel core i5 – 10th gen
- GPU : GFX cuda tool
- SSD: 500 GB
- RAM: 16GB

## IV. SYSTEM FEATURES

- 1) User-friendly based GUI built using industrial standard PyQt5.
- 2) Real time American standard character detection based on gesture made by user.
- 3) Customized gesture generation.
- 4) Forming a stream of sentences based on the gesture made after a certain interval of time
- 5) Real time (ASL) detection based on gesture made by user.

## V. PROPOSED SYSTEM

- 1) In this system our application will be able to detect the gesture of sign language in real time and convert it into the text form for the user.
- 2) In addition to the feature of this application
- 3) Given text output will be converted into the audio format which makes the interpretation much easy for the ill-educated people.

## VI. IMPLEMENTATION

- 1) Data collection: Collecting a large dataset of images or videos of individuals performing various sign gestures.
- 2) Data Preprocessing : Preprocessing the dataset involves standardizing the image size, normalizing the pixel values, and augmenting the dataset by adding variations to the images such as rotation, translation, and flipping to increase the robustness of the model.
- 3) Model Architecture: Convolutional, pooling, and fully connected layers are all part of the CNN model architecture design process. Pooling layers reduce the size of the feature maps, while convolutional layers extract features from the input images. The classification process is carried out by the fully connected layers.
- 4) Training: The CNN model is trained by feeding the preprocessed dataset into the model and iteratively adjusting the model's weights and biases to minimize the loss function.
- 5) Testing: Evaluating the performance of the trained model on a separate test dataset.

## VII. SYSTEM FLOW

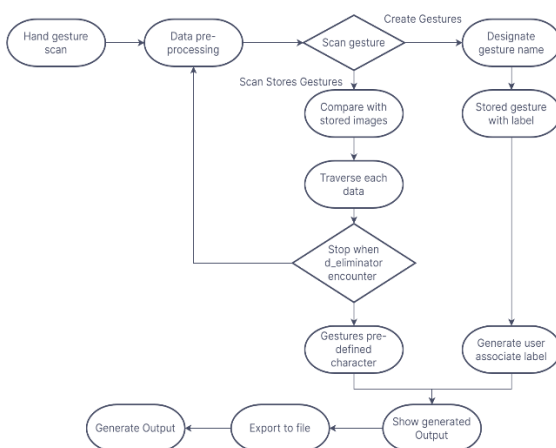


Fig. 1.1

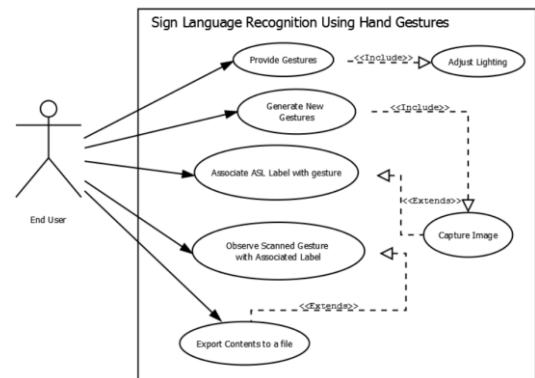


Fig. 1.2

## VIII. ALGORITHM

### CNN (Convolutional Neural Network) :-

A Deep Learning technique specifically created for working with images and videos is the convolutional neural network. It uses photographs as inputs, extracts and learns the image's attributes, then categories the images using the learnt features. This program takes its cues from how the Visual Cortex functions in the human brain. Processing of visual data from the outside world is carried out by the visual cortex, a region of the human brain. It comprises a number of layers, and each layer functions independently, extracting different information from images or other visuals. Once all the information from each layer has been combined, the picture or visual is then evaluated or classed step.

### Technology Used :-

#### TensorFlow :-

Google's open source TensorFlow artificial intelligence framework enables high performance numerical computing and machine learning. Dataflow graphs are created and executed using C++ using the TensorFlow Python framework. Many classification and regression techniques as well as, more generally, deep learning and neural networks are supported.

#### Keras API :-

Keras is a deep learning API built on top of the TensorFlow machine learning framework and based on Python. It was designed to make it easier to conduct quick experiments. To be successful in research, one must be able to move quickly from a concept to a conclusion.

#### OpenCV :-

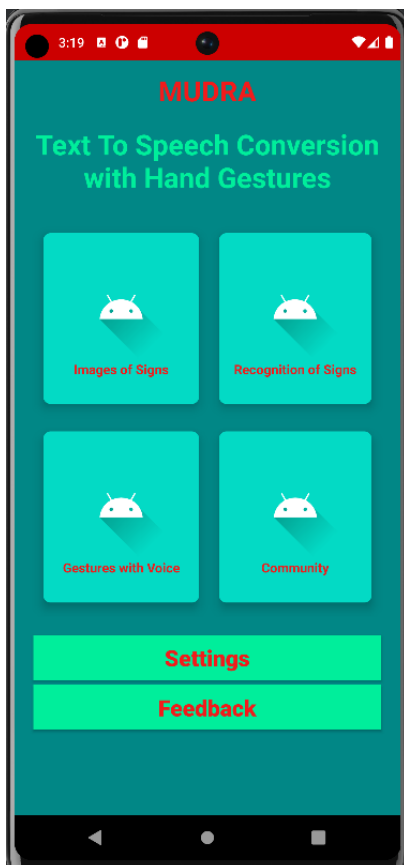
OpenCV-Python is a collection of Python bindings designed to address computer vision issues. Python is a universally useful programming language made by

Guido van Rossum that immediately acquired notoriety, generally because of its usability and meaningful code.

## IX. APPLICATIONS

- 1) Human-Computer Interaction: Sign gesture recognition can be used to interact with computers and other devices using sign language. This can help individuals with disabilities or those who prefer using sign language over traditional input methods.
- 2) Healthcare: Sign gesture recognition can be used to monitor the physical therapy progress of patients who are recovering from injuries or disabilities that affect their ability to perform sign gestures.

## X. RESULT



## XI. FUTURE SCOPE

- 1) It can be combined with a number of communication and search engines, including Google and What-App. In order for even those who are illiterate to be able to communicate with others or search the web with only gestures.
- 2) This project is now focusing on images, but with more advancement, it may be possible to use TTS to assign motion in a video sequence to a coherent speech. A video calling system can integrate the model with text to speech. In this way, the user can make motions, and the call recipient will be able to read or hear the message. The message will be sent to the hearing/speech challenged user while the receiver responds (subtitles).

## **XII. CONCLUSION**

Today's applications require different types of images as data sources for explanations and analyses. A number of functions must be implemented to perform different tasks.

Image degradation occurs when an image is converted from one format to another, such as sharing, scanning, digitizing, storing, etc. Thus, the output image must undergo a process called image enhancement, which involves a number of approaches designed to develop the visual presence of the image. Image enhancement improves the interpretability or awareness of image information for human listeners, while making excellent contributions to other autonomous image processing systems. The image is then extracted with different features using different approaches to make the image more computer readable. A powerful tool for expert data preparation, edge detection and combining error data from multiple sources. The goal of CNN is to get the correct classification.

## **XIII. REFERENCES**

- “Real-time Sign Language Recognition using Computer Vision” (2021) IEEE
- “Android based American Sign Language Recognition System with Skin Segmentation and SVM” (2018) IEEE
- “Sign Language Recognition for Speech and Hearing Impaired by Image Processing in MATLAB” (2018) IEEE
- “Real-Time Recognition of Indian Sign Language” (2019) IEEE