Sign Language Detection

Suhruth R [1], Sourav Nagesh [1], Vishwas H S [1], Sai Nagesh C H [1], Mrs.Manasa Sandeep[2]

[1] Student, Department of CSE, Dayananda Sagar Academy of Technology and Management, Bangalore, India

[2] Professor, Department of CSE, Dayananda Sagar Academy of Technology and Management, Bangalore, India

Abstract-Deaf and mute individuals, who make up approximately 5% of the global population, often rely on sign language to communicate with others. However, many of them may not have access to sign language, causing them to feel disconnected from others. To address this communication gap, a prototype for an assistive medium has been designed that allows individuals to communicate using hand gestures to recognize different characters, which are then converted to text in real-time. This system utilizes various image processing techniques and deep learning models for gesture recognition. Hand gestures have the potential to facilitate human-machine interaction and are an essential part of vision-based gesture recognition technology. The system involves tracking, segmentation, gesture acquisition, feature extraction, gesture recognition, and text conversion, all of which are critical steps in the design process. Overall, this technology has the potential to help bridge the communication gap between deaf and mute individuals and those who can hear and speak.

Key Words: OpenCV, Python, facial recognition, LSTM, SVM,RNN, ANN.

1.INTRODUCTION

Sign language is a method of communication that heavily relies on hand movements and facial expressions. While people with hearing impairments often use it to communicate with each other, those without hearing impairments rarely do so, which limits their social interactions with the hearing-impaired community. While real-timetranslation with interpreters is an option, it may not always be feasible and can also be expensive. Thus, an automatic translation system could prove beneficial in bridging this communication gap. In recent times, many strategies have been developed in this area toaid in the translation of sign language. This project focuses on developing a program for translating sign language into OpenCV and proposes a method of recognizing and translating personal signlanguage into standard text. Sign languages, also known as sign languages, are complete natural languages that use visual cues to convey meaning. Sign languages use manual communication, which involves the use of hand gestures, facial expressions, andother body movements, and have their own grammar and lexicon. Although there are some similarities between different signlanguages, they are not universal or fully understood. Linguistsconsider both spoken and sign languages to be natural forms of language that evolved over time without any conscious planning. Itis essential to note that sign language is not limited to people withhearing or speech impairments. Instead, it is often considered aprominent means of communication, particularly in situationswhere noise or distance hinders spoken communication.

While body language is a form of non-verbal communication, it should not be confused with sign language. Sign language uses manual communication to convey meaning and does not rely on acoustically transmitted sound patterns. Instead, it uses hand gestures, facial expressions, and other body movements such aseye and leg movements to convey meaning.

This paper proposes a design for character recognition and interpretation that can help in overcoming some of the communication barriers between people who use sign language and those who do not. Some of the challenges that people with hearing or speech impairments face when communicating withpeople who do not use sign language include social interaction, communication dissonance, education, behavioral problems, mental health, and safety concerns.

Gestures can be considered physical actions performed by the hand,eye, or any other part of the body, and hand gestures are the most easily interpretable for humans. The proposed mark recognition system can recognize marks with high accuracy and at a lower cost of features and time. By bridging the gap between people who use sign language and those who do not, this system can help promote inclusivity and enable better communication for all.

2.MOTIVATION

The motivation behind the "Sign Language Detection Using LSTM" project is to address the communication gap between people who use sign language and those who do not. Sign language is the primary means of communication for many deaf or hard of hearing individuals, but it is not widely understood, which creates significant challenges in effective communication and inclusion. Therefore, the project aims to develop a system that can accurately detect and interpret sign language gestures, making it easier for people who rely on sign language to communicate with others who do not.

In addition to increasing inclusivity and accessibility for people with hearing impairments, the project aims to empower individualswho use sign language by enabling them to communicate independently. By having a reliable sign language detection systemat their disposal, these individuals can gain more autonomy in theirdaily lives. They can use the system to express their thoughts, needs, and emotions without relying solely on interpreters or written text. This empowerment can significantly increase their self-confidence, self-expression, and overall quality of life.

Real-time communication is another key motivation for the project. The ability to interpret sign language gestures in real-time can be invaluable, especially in situations where instant communication is essential. For example, in emergency situations, medical facilities, or when interacting with the public, a system that can quickly and accurately interpret sign language can facilitate efficient and effective communication. This can help ensure early help, access to vital services, and equal opportunities for individuals who rely on sign language.

The project also seeks to support the teaching and learning of sign language. With the development of a sign language detection system using LSTM, it can serve as a valuable tool to help individuals who are learning sign language. The system can provide visual feedback, guidance, and assistance in understandingand practicing various sign gestures. It can be integrated into educational platforms, interactive applications, or virtual tutors to enhance the learning experience, promote consistency, and providepersonalized feedback to students.

Finally, the project is in line with the advancement of technology and contributes to the field of deep learning and pattern recognition. Developing a sign language detection system using LSTM involves solving complex problems such as analyzing time dependencies and recognizing subtle hand movements. By pushing the boundaries of research and innovation in computer vision and machine learning, the project can contribute to the development of more robust and accurate systems for a variety of real-world applications. In summary, the project aims to provide an effective means of communication for people who use sign language, promote inclusivity and accessibility, empower individuals, support teaching and learning, and advance technology.

3.PROBLEM STATEMENT

Sign language is a primary means of communication for many deaf or hard of hearing individuals, but the lack of widespread knowledge and understanding of sign language creates significant communication challenges. The available solutions, such as human interpreters or written text communication, have limitations in terms of availability, accessibility, and real-time responsiveness, leading to a lack of inclusivity and independence for sign language users. Moreover, sign language involves complex hand movements, facial expressions, and body language that require accurate and timely interpretation for effective communication. Therefore, this project aims to develop a sign language detection system using LSTM, a deep learning architecture that can efficiently model sequential data. This system seeks to accurately recognize and interpret sign language gestures from video inputs in real-time, enabling effective communication between sign language usersand non-sign language users. The ultimate goal of this project is to bridge the communication gap, promote inclusivity, empower individuals who use sign language, and enhance their independent communication in various areas such as education, employment, healthcare, and social interactions.

To achieve this goal, the project needs to address several challenges such as effectively capturing and representing temporal dependencies in sign language gestures, handling the variability and nuances of different sign language styles and contexts, ensuring real-time response for seamless communication, and achieving high levels of accuracy and robustness in gesture recognition. The development of a sign language detection system using LSTM would contribute to a more inclusive society where individuals relying on sign language can communicate effectively and participate fully in various activities, reducing barriers and promoting equal opportunities for all.

4.LITERATURE SURVEY

A method for recognising human-computer gestures was createdby Marouane Benmoussa and colleagues. Recent years have seen agreat deal of research in gesture recognition, particularly as human computer interaction (HCI) technology has improved. This final point can be improved by teaching computers how to communicate in human ways. The objective is to enable computers tocomprehend human speech, as well as facial expressions and gestures. Using Kinect's Skeletal Tracking, the system can identify hand motions, follow them, and decipher their meanings. It employs Scale Invariant Features Transform (SIFT) and Speeded Up Robust Features (SURF), both of which were trained using K- means and Support Vector Machine (SVM) classifiers.We first collected depth photos for 16 different movements using the Microsoft Kinect sensor in order to build our HGR. Using depth data can clear up any uncertainty relating to the background. The keypoint sets produced by SIFT and SURF functions can be usedto represent the hand gesture training images, but the amount of keypoints from each image varies and lacks a clear hierarchy. To resolve this issue, we employ a bag of words strategy. One of the most well-known approaches in computer vision is bag-of-words (Bow). Additionally, by a difference of 50 keypoints, SURF was able to define the gesture keypoint better than SIFT.The effectiveness of SURF against scale, rotation, and translation biases has also been demonstrated. In this study, we presented a machine learning technique to instantly identify 16 user hand motions using the Kinect sensor. A support vector machine model is trained using hand depth data, and a bag of words including SIFT and SURF descriptors is then extracted. By measuring the method's performance using the area under the ROC curve, itproduced an SURF performance of 98% and a SIFT performance of 91%. We can infer that SURF is more appropriate for real-time applications because it is three times faster than SIFT.[1]

A system for recognising human and computer hand gestures was suggested by Greeshma Pala et al. Technology that recognises hand gestures has grown in importance recently since it helps with communication and offers a natural way to communicate that can be applied to many different tasks.The user interface for this system has three options or modes at the outset. Hand Sign to Text Recognition, Hand Gesture to Speech, and Hand Gesture Recognition. We have chosen a vision-based method that does not require any additional technology to recognise hand motions. Using a web camera, pictures were gathered. According to the generated partition, images must be trained and tested. The image is first made into a grayscale version. To ensure uniformity across all pictures, each one is downsized to 75x75. Where p and q are two points in Euclidean n-space, q and p are Euclidean vectors starting from the space's origin (the starting point), and n is n space, the K-Nearest-Neighbor (KNN) classification algorithm is applied. The majority of the classes to which the points belong are used to categorise the picture field. Regression and classification are done using SVM. It is specifically employed to identify the ideal separating line. The creation of the ideal separation hyperplane is the main objective of SVM. The neurons that make up a CNN have parameters like weights and biases that can be learned.CNNs are distinguished by their explicit presumption that the entities are pictures. The pickle model is used to load the dataset. The train and test split is carried out after the data set has been loaded. This article investigates the most effective algorithm for hand sign recognition. examines how data augmentation affects deep learning. Then, with good accuracy and little loss, we may say that CNN is, on balance, the best experimented algorithm.[2]

A human-computer hand gesture recognition and classification system for the deaf and mute was proposed by Nitesh S et al. A gesture can be defined as any physical action made with the hand, eye, or any other part of the body. The most humane and simple to understand motions are hand gestures. The webcam input image is recorded and can be utilised as an input image for character recognition or to store as a training dataset. Images that are captured are RGB files. The photos that were obtained had very high pixel values and dimensions. Therefore, we use the "rgb2gray" function to convert the RGB image to grey and thenthe grey image to a binary image.The process of segmentation separates the image into two parts: the background and the foreground, which contains the area of interest. Principle Component Analysis (PCA) is used to extract features. To determine these eigenvalues and eigenvectors, all of the pictures are combined into a column matrix, which is then averaged and then subtracted for normalisation. Hand gestures are organisedaccording to Euclidean distance. For the data set, eigenvectors were constructed. The dataset is created and saved during thetraining phase, which is the first step. The accuracy of the system is directly inversely correlated to the number of photos stored per character. This matrix is used to calculate the eigenvector, which is utilised to obtain the element vector. The second phase, known as the test phase, calls for the recognition of an input gesture. The gesture is identified and a maximum score based on Euclidean distance is calculated. The percentage accuracy for each character is then displayed in the accuracy table. Character identification for the input gesture is facilitated by the Euclidean distance. Through this approach, persons with disabilities can interact directly and without sign language with everyday people. The system'saccuracy is influenced by the lighting.[3].

A deep learning-based character action detection system was proposed by Shivanarayna Dhulipala et al. Cognitive issues, such as those pertaining to sign languages and their constraints, have become easier to tackle as a result of the rapid rise in computer useand artificial intelligence. The most important development in artificial intelligence is deep learning, which is used to teach computer systems how to recognise, decipher, and translate letters into written language. As a result, the focus of this dissertation willbe on employing deep learning to apply LSTM and CNN models to the detection of human activity and sign language. to bridge the communication gap between the hearing and the deaf. A CNN model is a crucial component of a neural network used for character and face detection and identification that recognises and classifies images. The neurons that make up CNN models have biases and learnable weights. Specific neurons receive input data, and weighted sums are calculated that, in response to actions,activate specific functions and produce specific outputs. The multi-channel mode frequently employs CNN models.

5.TECHNOLOGIES USED

A. Python

Python is a high-level, general-purpose programming language thatis interpreted, meaning that it can execute code line-by-line without the need for compilation. One of the distinctive features of Python is its emphasis on code readability, achieved through the use of significant whitespace characters. The language's object- oriented programming paradigm and various language constructs make it suitable for developing both small and large projects, enabling developers to produce code that is clear and maintainable.

IDE (Jupyter)

Jupyter Notebook offers us a user-friendly, interactive data science environment for a variety of computer languages that serves as both an IDE and a tool for presentations or instruction.

Numpy (version 1.16.5)

The Python package NumPy is used to manipulate arrays. Additionally, it provides functions for working with matrices, the Fourier transform, and linear algebra's dominance. In the year2005, Travis Oliphant developed NumPy. Numerical Python isreferred to as NumPy.

B. OpenCV

OpenCV is a sizable open-source toolkit for image processing, machine learning, and computer vision. Numerous programming languages, including Java, C++, Python, etc., are supported by OpenCV. It can analyse movies and photos to recognise faces, objects, and even handwriting. The number of weapons in your arsenal expands when paired with numerous other libraries, such as Numpy, a highly developed library for numerical operations; any operations that can be performed in Numpy can also be combined with OpenCV. Using a vast array of OpenCV projects and programmes, the OpenCV Tutorial teaches you imageprocessing fundamentals to advanced techniques, including image and video operations.

C. Keras

High-level neural network library Keras utilises TensorFlow, CNTK, and Theano as its foundation. Using Keras for deep learning enables quick and simple prototyping as well as seamless CPU and GPU operation. This framework was created using the debugging-friendly and resilient Python programming language.

D. TensorFlow

A complete open-source machine learning platform is called TensorFlow. It's a vast and adaptable ecosystem of resources, frameworks, and tools that offers workflow with high-level APIs. You can choose from a variety of idea levels offered by the framework to create and implement machine learning models.

2. SYSTEM DESIGN

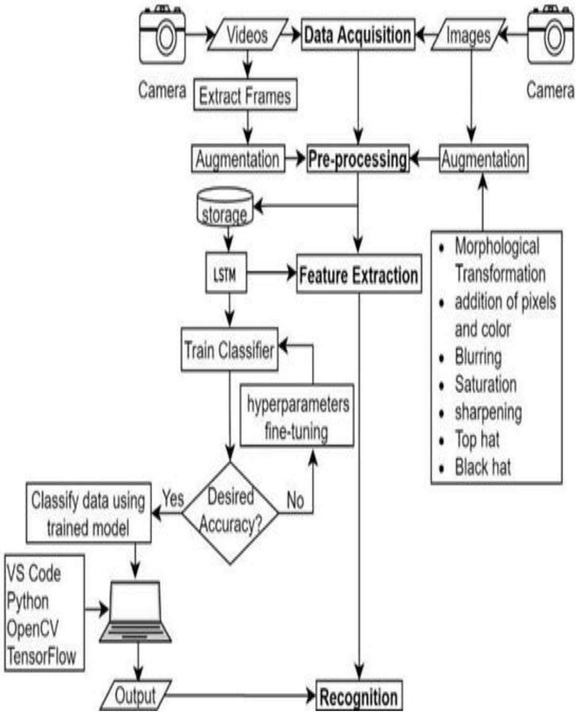


Fig-1: System Architecture

- DATA FLOW

Here, the image is subjected to a classifier that bases its decisions on preferences and, when motion is present, the trajectory itself. Keyframe and feature extraction come after that. The generation ofresults is then followed by the classifier based on forms in the key of the hand.

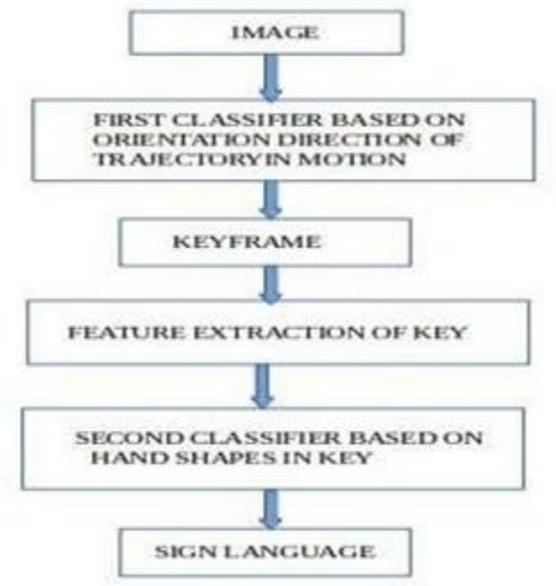


IMAGE
FIRST CLASSIFIER BASED ON ORIENTATION DIRECTION OF TRAJECTORYIN MOTION
KEYFRAME
FEATURE EXTRACTION OF KEY
SECOND CLASSIFIER BASED ON HAND SHAPES IN KEY
SIGN LANGUAGE

Fig-2: Data flow diagram

- USE CASE DIAGRAM

The user and the system are two actors in this scenario. They are given several chores to complete. The user is in charge of turning on the webcam, taking the picture, and receiving the results. The system handles anything else. Here, the user will activate the webcam and record his motion within the live video stream. The system will translate the gesture, extract its features, and compare them to features that already exist before performing hatching and gesture recognition. The user can understand the gesture's meaningafter the results are displayed.

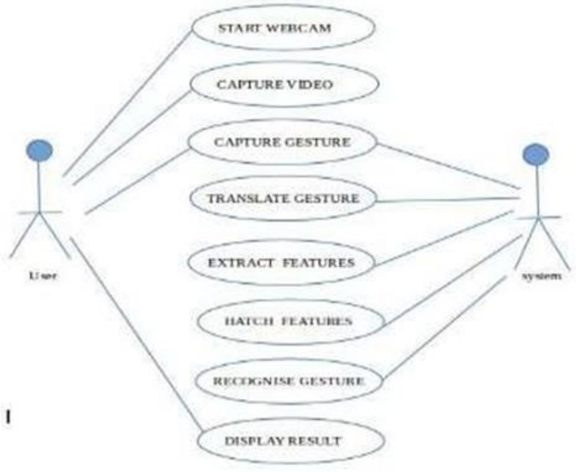


Fig-3: Use case diagram

- SEQUENCE DIAGRAM

It shows how several objects interact with one another in a sequential manner. That is the order in which these events happen. This sequence diagram is frequently referred to in event diagrams or event scenarios. This diagram illustrates the interactions and sequential order of the system's components. The sequence diagram pertinent to this project is shown below. Seven steps are required to arrive at the outcome in the diagram below. The first stage is for the user to record a video using a camera, which is followed by picture capture. The system recognises the hand in the third sequence, extracts the feature from it in the fourth, and maps the feature after dataset matching. The user is finally presented with results that match.

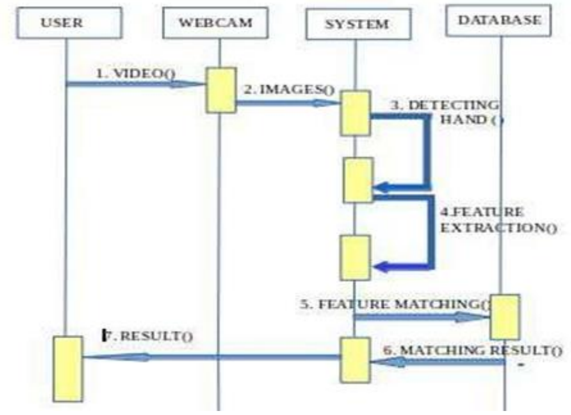


Fig-4: Sequence diagram

6.BASIC MODULES

A. Keypoints and extraction using MP Holistic Module One of the pipelines, Mediapipe Holistic, comprises enhanced face, hand, and pose components that enable holistic tracking. As a result, the model is able to simultaneously recognise hand and body poses as well as facial landmarks. Face and hand detection and keypoint extraction for feeding into a computer vision model are two of the primary uses of the comprehensive MediaPipe.

A comprehensive and multimodal applied machine learning pipeline can be built using MediaPipe, an open-source, cross- platform machine learning framework. Developers can concentrate more on experimenting with models than with the system by using MediaPipe to handle model implementations for systems runningon any platform.

A feature is used to access the image input from the system webcam using the OpenCV framework, hand and face landmark detection, and keypoint extraction.

B. Collect Keypoint Values for Training and Testing Module The MediaPipe Hand Landmarker task enables the detection ofhand landmarks in an image through the use of functions such as detect, detect_for_video, and detect_async. The process for handlandmark detection involves preprocessing input data, detectinghands in the image, and identifying hand landmarks. In order torun Hand Landmarker in video or live mode, a timestamp for the input image is required. When running in live mode, the currentthread is not blocked and results are returned immediately. Uponcompleting processing of an input frame, the Hand Landmarkertask invokes its result listener with the detection outcome.

The output of HandLandmarkerResult is composed of three arrays,where each element includes the results for one detected hand. These arrays comprise of:

Handedness: representing whether the detected hands are left or right-handed

Landmarks: consisting of 21 landmarks with x, y, and z coordinates. The x and y values are normalized between 0.0 and

1.0 based on the width and height of the image, while the z coordinate is relative to the wrist's depth. Smaller values indicate landmarks closer to the camera, and the z scale is roughly the sameas the x scale.

World Landmarks: similarly comprising of 21 landmarks with actual 3D coordinates represented by x, y, and z values in meters, originating from the geometric center of the hand.

C. Build and Train LSTM Neural Network Module Numerous sign language recognition (SLR) systems have been created by researchers, but they are only capable of recognising discrete sign motions. In this study, we propose a continuous SLR model that recognises a series of concatenated gestures as a modified long-short-term memory (LSTM) model for contiguous gesture sequences. Its foundation is the subdivision of continuous characters into smaller parts and the neural network modelling of those smaller units. As a result, during training, consideration of a distinct sub-unit combination is not necessary.

D. Real-Time Detection Module The live hand gestures are translated using this approach into letters, then words, and finally sentences.The process of doing sign detection in real-time with quick inference and a minimal level of accuracy is known as real-time sign detection. According to one definition, a real-time system "controls an environment by receiving data, processing them, and returning the results in a timely manner to have an immediate impact on the environment."

7.IMPLEMENTATION

- We start by collecting keypoints from mediapipe holistic and collect lots of data from keypoints i.e. from our hands, on our body and on our face and store the data in the form of numpy arrays. We can change the number of sequences according to our needs, but each sequence will have 30 frames.
- Then we create an LSTM model and train with our stored data, which helps us detect action with a series of frames.
- The number of epochs for the model is determined by us, if we increase the number of epochs, the accuracy will increase, but it will also increase the time needed to run the model and the gesture recognition model may be reassembled.
- After completing the training, we can use this model to detect hand gestures in real time and simultaneously convert them to text using OpenCV.

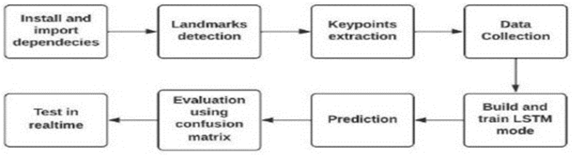
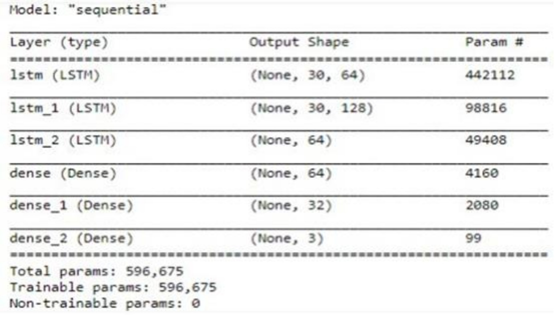

Install and import dependecies
Landmarks detection
Keypoints extraction
Data Collection
Build and train LSTM mode
Prediction
Evaluation using confusion matrix
Test in realtime

Fig-5: Methodology



```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 30, 64)            442112
_________________________________________________________________
lstm_1 (LSTM)                (None, 30, 128)           98816
_________________________________________________________________
lstm_2 (LSTM)                (None, 64)                49408
_________________________________________________________________
dense (Dense)                (None, 64)                4160
_________________________________________________________________
dense_1 (Dense)              (None, 32)                2080
_________________________________________________________________
dense_2 (Dense)              (None, 3)                 99
=================================================================
Total params: 596,675
Trainable params: 596,675
Non-trainable params: 0
```

Fig-6: Model Summary

8.RESULTS

The objective of this study was to use forearm, hand, and finger kinematics models along with deep neural networks and Mediapipe Holistic to predict signals. The Mediapipe LSTM with data augmentation produced the greatest results, averaging 91.1 percent accuracy on the test sets. This sign language detector will be able to comprehend signals as well as recognise and detect hand and produce coordinators. Real-time updates will be made to all of the signs.

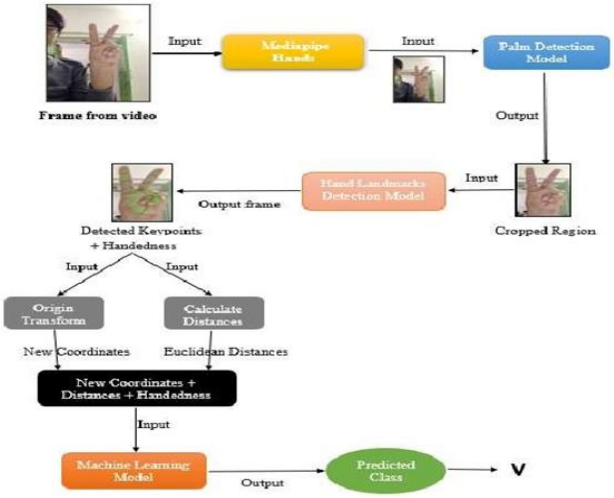


Fig-7: Real-time Workflow of Project

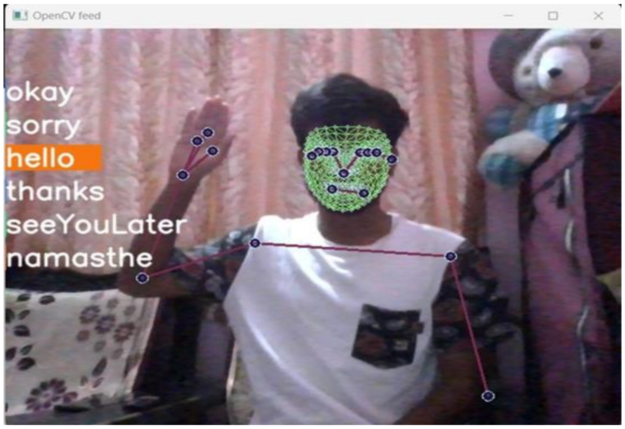


Fig-8: Detecting the sign for Hello

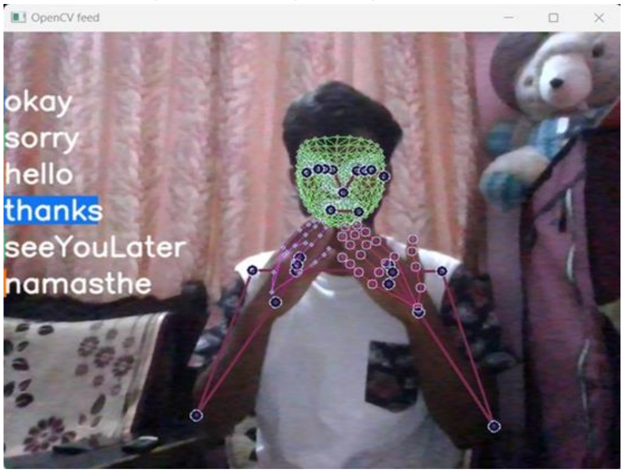


Fig-9: Detecting the sign for Thanks

9.LIMITATIONS

Although our project "Sign Language Detection" offers significant potential, it also has several limitations that need to be considered:

- Size and diversity of the training data set: The quality, size, and diversity of the training data set have a significant impact on the performance of an LSTM model. Working with siglanguage users to ensure that all sign language styles, dialects, and situations are represented while also gathering a big and varied set of sign language video data can be difficult. Reduced generalisation and accuracy of the model may result from the data set's small size and lack of diversity.
- Gestures in sign language come in a range of shapes and sizes,each with its own variants and complexities. An LSTM model may find it difficult to distinguish minute changes between comparable movements or to accurately record minute details. Dealing with the diversity and subtleties of many sign language dialects, styles, and individual differences is a difficult task that could compromise the accuracy anddependability of the system.
- Latency and real-time processing: To ensure seamless and prompt communication, real-time sign language identification calls for quick and efficient processing. The computing demands of LSTM models can be high, which may cause latency problems. Accuracy and real-time performance can be difficult to achieve, especially on devices with limited resources like embedded systems or smartphones.
- An LSTM model built on a particular dataset could have trouble generalising to new users or surroundings that weren't encountered during training. Everybody signs differently, and environmental elements like lighting, camera angles, and background clutter can interfere with gesture detection. To ensure optimal performance, the system can need additional tweaking or adaption to new users and circumstances.
- Accessibility and user interface design must be taken intoaccount when deploying a sign language detecting system in practical situations. To ensure the system's widespread applicability and acceptance, it should be user-friendly, intuitive, and adaptable to many devices or platforms. It is crucial to address the difficulty of ensuring accessibility for people with various levels of technical knowledge and disabilities.

While the project has these limitations, they can be mitigated through ongoing research, continuous improvement, and collaboration with sign language users and experts to refine the system's accuracy, robustness, and usability in real-world settings.

10.CONCLUSION AND FUTURE ENHANCEMENT

Deaf and mute individuals face numerous obstacles in their daily lives and are unable to lead normal lives as basic facilities are not available to them in India. They must rely on family members or interpreters to travel in public transport or attend school or work. The proposed system is a real-time sign language translator that can assist them in communicating with normal people, allowing them to engage in various activities and be treated like any other normal person. The system receives input gestures through the camera and displays the word according to the trained data, although the accuracy of the system is sometimes affected bylighting conditions. Our work aimed to develop an automatic real- time sign language gesture recognition system using various tools, but there is still room for further improvements.

Future studies could potentially develop a web or mobile application that can classify complete word symbols using facial emotions and relative hand movements from the face, which could be available on Android and Apple platforms. The future of sign language detection lies in the use of LSTM. With the increasing availability of sensors and wearable technology, there is growing interest in developing sign language recognition systems that can be used in real-world applications. For instance, such systemscould be used to facilitate communication between deaf or hard of hearing individuals and those who do not know sign language.

REFERENCE

[1] Machine Learning for Hand Gesture Recognition Using Bag-of-words Marouane Benmoussa, Abdelhak Mahmoudi, LIMIARF, Ecole Normale Superieure,Mohammed V University, Rabat, Morocco, 2018.

[2] Machine Learning-based Hand Sign Recognition Ms.Greeshma Pala, Ms.Jagruti Bhagwan Jethwani, Mr.Satish Shivaji Kumbhar,M s. Shruti Dilip Patil,Department of ComputerEngineering and Information Technology, College of EngineeringPune,

Pune, 2021.

[3] Online Hand Gesture Recognition & Classification for Deaf & Dumb Nitesh S.Soni, Prof.Dr.M.S.Nagmode, Mr.R.D.Komati, Department of Electronics and Telecommunication, MIT College of Engineering, Pune, 2016.

[4] Sign and Human Action Detection Using Deep Learning Shivanarayna Dhulipala, Festus Fatai Adedoyin and Alessandro Bruno, Department of Computing and Informatics, Bournemouth University, Poole, 2022.

[5] Sign language Recognition Using Machine Learning Algorithm Radha S. Shirbhate, Vedant D. Shinde, Sanam A. Metkari, Pooja U. Borkar, Mayuri A. Khandge, JSPM's BSIOTR – Wagholi, Pune, IRJET, 2020