www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

HAND SING DETECTION(ASL) USING AI AND IMAGE PROCESSING

[1]Dr.S.K.UMAR FARUQ,[2]B.JAMUNA,[3]A.DEEPIKA,[4]B.VIJAY,[5]A.VENKAT SAI,[6]A.YESHWANTH

[1](PROFESSOR),HOD,ECE, TEEGALA KRISHNA REDDY ENGINEERING COLLEGE.

[2](ASSISTANT PROFESSOR),ECE, TEEGALA KRISHNA REDDY ENGINEERING COLLEGE.

[3456]B.Tech. Scholar, ECE, TEEGALA KRISHNA REDDY ENGINEERING COLLEGE.

ABSTRACT

The increasing need for accessible communication for the hearing-impaired community has led to advancements in technology, particularly in the field of Hand Sign Detection for American Sign Language (ASL). This project explores the development of an AI-driven hand sign detection system using image processing techniques in Python. By leveraging convolutional neural networks (CNNs) and machine learning algorithms, the system is capable of recognizing and interpreting ASL gestures from live video streams or static images. The model is trained on a dataset of ASL hand signs, using Python libraries such as OpenCV for image preprocessing and TensorFlow or Keras for building and training the neural network. The system processes input images, identifies hand gestures, and maps them to their corresponding ASL letters or words, providing real-time feedback. This project aims to bridge communication gaps by offering a tool that can be used for learning ASL or assisting in daily interactions between the hearing and hearing-impaired communities. The proposed system has applications in educational tools, assistive technologies, and real-time translation services. With further training and optimization, it has the potential to improve the quality of life for individuals who rely on sign language as their primary mode of communication.

Index Terms: American Sign Language (ASL), Hand Sign Detection, Convolutional Neural Networks (CNN), Image Processing, Real-time Translation, Assistive Technology, Machine Learning, Python, OpenCV, Keras, TensorFlow, Human-Computer Interaction.

1.INTRODUCTION

A gesture is any movement of a body part, such as the face or the hand. Image processing and computer vision are used here for sign language recognition and python text to speech is used for speech conversion. Sign recognition allows computers to understand human actions and serves as a translator between computers and humans. This could allow humans to engage naturally with computers without coming into direct contact with the mechanical equipment. Deaf and dumb people use hand and pose gestures to communicate in sign language. When transmitting voice is impossible or typing and writing is problematic, but there is the possibility of seeing, this community uses sign language to communicate. The sole means of communication between people at the moment is sign language.

Normally, everyone uses sign language when they don't want to speak, but for the deaf and dumb community, this is their sole means of communication. The same meaning is conveyed through sign language as it is through spoken language. All around the world, the deaf and dumb community uses this, though in localized forms like ISL and ASL. One or two hands can be used to make hand gestures when utilizing sign language. There are two types of it: continuous sign language and isolated sign language. Continuous ISL, also known as Continuous Sign language, is a series of movements that produce a coherent sentence as opposed to isolated sign language, which consists of a single motion with a single word. We used an independent ASL gesture recognition algorithm in this work.

Humans can communicate with one another in a variety of ways. This include behaviour such as physical gestures, facial expressions, spoken words, etc. However, those who have hearing loss are restricted to using hand gestures to communicate. People with hearing loss and/or speech impairments communicate using a standard sign language that is incomprehensible to non-users.

Sign language is the communication system for those who are hard of hearing and deaf. It ranks as the sixth most utilized language worldwide. It is a type of communication that uses hand movements to communicate ideas. Each region has its specific sign language like normal language. In 2005 there were an estimated 62 million deaf people worldwide and about 200 different sign languages in use around the world, many of which have distinctive features.

ASL is the primary language of many deaf citizens in North America. Hard-of-hearing and hearing people also use it. Hand gestures and facial expressions are used to convey this language. The deaf community has access to ASL as a means of communication with the outside world and inside the community. But not everyone is familiar with the signs and motions used in sign language. Understanding sign language and being familiar with its motions takes a lot of practice. Since there are no reliable, portable tools for identifying sign language, learning sign language takes a lot of time. However, since the development of neural networks and deep learning, it is now possible to create a system that can identify things, or even objects of different categories.

Fig.1.1-American sign languages.

OBJECTIVES

- According to statistics, over 80% of specially abled individuals are illiterate, and the system tries to bridge the gap between a normal, a hearing-impaired and a visually impaired person by turning a majority of sign language to text and speech. People who are deaf or hard of hearing can communicate their message using gestures that can be read.
- People who are not visually challenged can use the software to comprehend sign language and communicate effectively with those who are. Also, people who are visually impaired can also communicate when the sign language predicted text is converted to speech.This project will bridge the gap of difficulty in understanding sign language that existed previously.
- To train a model, it will employ cutting-edge deep learning algorithms. The model will collect frames for gestures using the camera, train the model, and evaluate the precision for each gesture. The gesture will then be predicted in real time. The gesture is then translated to text and speech.

2.RELATED WORK

Previous researchers have emphasised their work on the prediction of sign language gestures to support people with hearing impairments using advanced technologies with artificial intelligence algorithms.

Although much research has been conducted for SLR, there are still limitations and improvements that need to be addressed to improve the hard-of-hearing community. This section presents a brief literature review of recent studies on SLR using sensor and vision-based deep learning techniques.

Literature review of the problem shows that there have been several approaches to address the issue of gesture recognition in video using several different methods. In the authors used Hidden Markov Models (HMM) to recognize facial expressions from video sequences combined with Bayesian Network Classifiers and Gaussian Tree Augmented Naive Bayes Classifiers.

Francois et al. also published a paper on Human Posture Recognition in a Video Sequence using methods based on 2D and 3D appearance. The work mentions using PCA to recognize silhouettes from a static camera and then using 3D to model posture for recognition. This approach has the drawback of having intermediary gestures which may lead to ambiguity in training and therefore a lower accuracy in prediction.

Let's approach the analysis of video segments using Neural Networks which involves extracting visual information in the form of feature vectors. Neural Networks do face issues such as tracking of hands, segmentation of subjects from the background and environment, illumination of variation, occlusion, movements and position. The paper by Nandy et al. splits the dataset into segments extracts features and classifies using Euclidean Distance and K-Nearest Neighbors

Similar work by Kumud et al. defines how to do Continuous Indian Sign Language Recognition. The paper proposes frame extraction from video data, pre-processing the data, extracting key frames from the data followed by extracting other features, recognition and finally optimization. Pre-processing is done by converting the video to a sequence of RGB frames. Each frame having the same dimensions. Skin colour segmentation used to extract skin regions, with the help of HSV. The images obtained were converted to binary form. The key frames were extracted by calculating a gradient between the frames. And the features were extracted from the key frames using oriental histogram. Classification was done by Euclidean Distance, Manhattan Distance, Chess Board Distance and Mahalanobis Distance. Each frame having the same dimensions. Skin colour segmentation used to extract skin regions, with the help of HSV. The images obtained were converted to binary form. The key frames were extracted by calculating a gradient between the frames. And the features were extracted from the key frames using oriental histogram.

3.SYSTEM ANALYSIS

3.1 Existing System:

Image processing to detect American Sign Language (ASL) hand signs.It employs Convolutional Neural Networks (CNNs) trained on ASL datasets for gesture recognition.OpenCV handles image preprocessing, while TensorFlow/Keras is used for model training and inference.The system provides real-time feedback from live video or static images to translate ASL gestures.

3.2 Proposed System

The proposed work considers the issues faced by prior models and works to reduce them. We created a system that does not compromise on efficiency or performance. The non-uniform background and segmentation of hands from the background was one of the key issues that researchers confronted. We solved that problem by creating a dataset using Google's Mediapipe solution and the openCV library to detect landmarks from the hand, which we also used during the real-time detection part of the hand gesture recognition system, so that regardless of whether one is in their car, at home,or on the street, the system will detect accurate landmarks from their hands.the suggested system consists of three primary modules that are linked in series. The Dataset module is where the landmarks are extracted and the dataset construction process is completed, followed by the Preprocessing module, where the data is processed to input into the final LSTM module, where the training for detecting gestures takes place.

4.SYSTEM ARCHITECTURE

The proposed system follows a vision-driven methodology to perform gesture-based sign recognition from frames extracted from video inputs. The sign recognition process consists of three phases, namely data collection, data pre-processing and feature extraction and gesture recognition. After the data collected is pre-processed and augmented , the feature extraction process is initiated. In this, facial features, landmarks of both hands, and bodily postures are extracted as keypoints from a sequence of input frames captured via a web camera. Then the extracted essential data points are considered as crucial features to be fed to the implemented classifier that recognizes the gestures performed by the user These recognized gestures and moments are further converted to the textual form and displayed on the screen in real time. The three phases of the system are discussed in detail below:

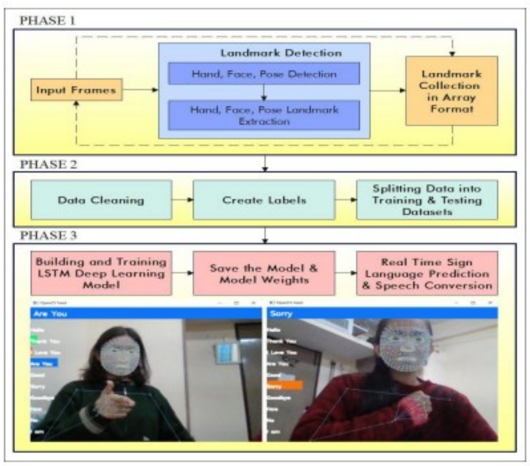

Fig.4.1-Architectural diagram of the system

5.METHODOLOGY

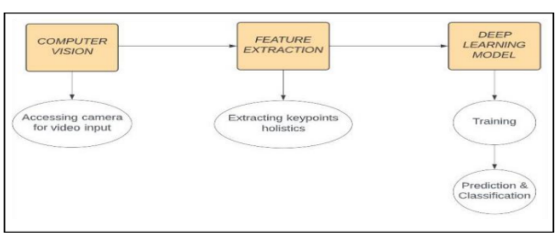

Fig.5.1 Methodology

Stage 1: Data Preprocessing and Feature Extraction

For data preprocessing and feature extraction from images, we applied a multistage pipeline using MediaPipe Holistic. For each input frame from the webcam, MediaPipe Holistic processed individual models for the hands, face, and pose components using region-appropriate image resolutions. The workflow of Stage 1 is briefly described below:

The human pose and subsequent landmark models were estimated using BlazePose's pose detector. Based on the inferred pose landmarks, three Region of Interest (ROI) crops—one for the face and two for the hands—were derived, and a re-crop step was performed to improve the ROIs.

Next, the corresponding landmarks were estimated. To achieve this, full-resolution input coordinates were cropped to the ROIs for task-specific hand and face models.

Finally, all landmarks were combined to yield more than 540 landmarks.

Stage 2: Data Cleaning and Labelling

After Stage 1, the landmark points were flattened, concatenated, and stored in a file. Null entries were then identified and removed. Data cleaning is essential to avoid failed feature detection, which can occur when a blurry image is submitted to the detector, resulting in null entries in the dataset. Training with such noisy data may reduce prediction accuracy and introduce

ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

bias. Labels were created for each class, and the associated frame sequences were saved to prepare the data for subsequent stages of training, testing, and validation.

Stage 3: Gesture Recognition and Speech Translation

With the preprocessed data, we trained an LSTM model to detect actions using a limited number of frames. The number of training epochs was selected carefully, as more epochs can improve performance but may also increase training time and risk overfitting.

Once the model is trained, it can recognize sign language in real time using the OpenCV module. The recognized text is then converted into speech using the Python text-to-speech module. The resulting audio file is played simultaneously using the system's default audio output device.

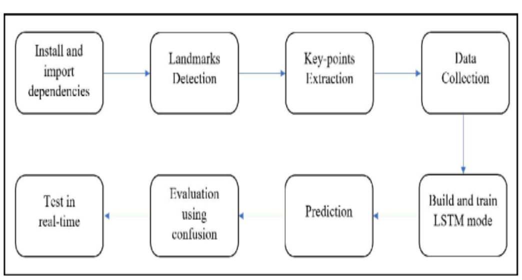

Fig.5.2 - Detailed Flow of Project.

6.RESULT

Fig.6.1-Capture Image

Fig.6.2-segmentation Part

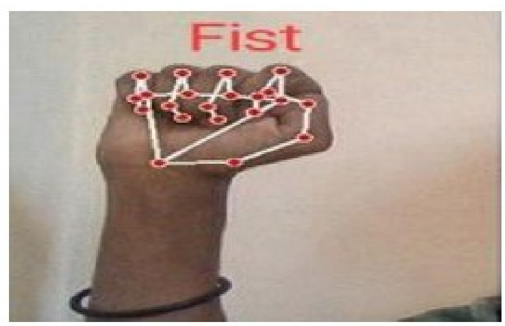

Fig.6.3-output of stage-1

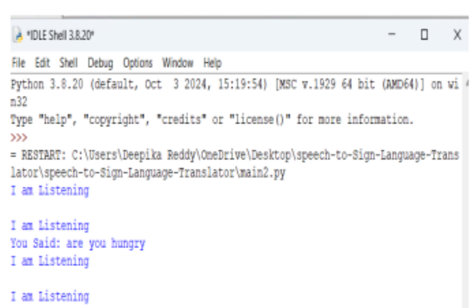

Fig.6.4-Taking input

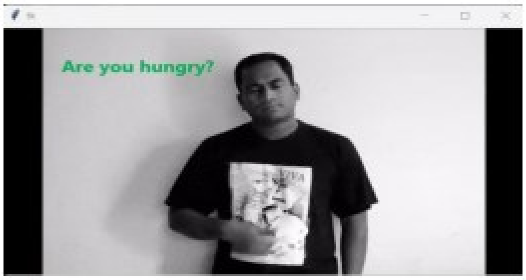

Fig.6.5-Responding in text format

7.CONCLUSION

Using the popular MediaPipe framework and LSTM networks, this comprehensive study proposed and developed a system for American Sign Language (ASL) recognition. Initially, a folder structure was created for each gesture, with 30 subfolders per gesture. These subfolders can be considered as video folders, each containing 30 frames represented as NumPy arrays. These arrays store landmark values detected and extracted using the MediaPipe Holistic solution.The collected data was then used to train an LSTM network, which achieved an accuracy of 80% on the testing dataset. Finally, the system was evaluated using real-time input fed directly into the model. The recognized gestures were displayed on the screen and translated into speech.Some lag was observed during real-time gesture recognition. Through this project, we learned that sometimes simple approaches can outperform more complex ones. Additionally, we gained insight into the challenges and time constraints involved in creating a dataset from scratch.

8.FUTURE SCOPE

- A complete product can be developed to assist speech- and hearing-impaired individuals, thereby reducing the communication gap. This can be achieved by putting the entire system online, allowing users to utilize their cameras to quickly recognize gestures.
- The system can be enhanced by incorporating a wider range of dynamic gestures that are commonly used. Currently, the model is trained on a limited number of words; it can be expanded to recognize full sentences, alphabets, and numbers. Training the model with data representing various skin tones, hand postures, lighting conditions, and environments will improve its robustness and accuracy.
- Integration with mobile devices can be explored by building a real-time mobile application. This would

enable users to recognize gestures on the go and allow them to select the language to be recognized as needed.

9.REFERENCES

1.K. Cheng, “Top 10 & 25 American Sign Language Signs for Beginners – The Most Known Top 10 & 25 ASL Signs to Learn First,” Start ASL, 29-Sep-2021. [Online]. Available: https://www.startasl.com/top-10-25-american-sign-language-signs-for-beginners/

2.A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian, and B. B. Chaudhuri, “A Modified LSTM Model for Continuous Sign Language Recognition Using Leap Motion,” IEEE Sensors Journal, vol. 19, no. 16, pp. 7056–7063, Aug. 2019.

3.V. Sharma, M. Jaiswal, A. Sharma, S. Saini, and R. Tomar, “Dynamic Two Hand Gesture Recognition using CNN-LSTM based Networks,” 2021 IEEE International Symposium on Smart Electronic Systems (iSES), 2021, pp. 224–229, doi: 10.1109/iSES53667.2021.00057.

4.K. Amrutha and P. Prabu, “ML Based Sign Language Recognition System,” 2021 International Conference on Innovative Trends in Information Technology (ICITIIT), 2021, pp. 1–6, doi: 10.1109/ICITIIT51526.2021.9399594.

5.M. Wurangian, “American Sign Language Alphabet Recognition,” Medium, 15-Mar-2021. [Online]. Available: https://medium.com/mlearning-ai/american-sign-language-alphabet-recognition-4dba892fbdcc

6.A. Dennisan, “American Sign Language Alphabet Recognition Using Deep Learning,” arXiv preprint, 10-Feb-2022. [Online]. Available: https://arxiv.org/abs/2202.04174

7.P. Likhar, N. K. Bhagat, and R. G. N., “Deep Learning Methods for Indian Sign Language Recognition,” 2020 IEEE 10th International Conference on Consumer Electronics (ICCE-Berlin), 2020, pp. 1–6.

8.K. Bantupalli and Y. Xie, “American Sign Language Recognition Using Deep Learning and Computer Vision,” 2018 IEEE International Conference on Big Data (BigData), 2018, pp. 4896–4899, doi: 10.1109/BigData.2018.8622141.

9.S. Shivashankara and S. Srinath, “American Sign Language Recognition System: An Optimal Approach,” International Journal of Image, Graphics and Signal Processing, vol. 11, no. 8, pp. 18–25, 2018.

10.N. K. Bhagat, Y. Vishnusai, and G. N. Rathna, “Indian Sign Language Gesture Recognition Using Image Processing and Deep Learning,” 2019 Digital Image Computing: Techniques and Applications (DICTA), doi: 10.1109/DICTA47822.2019.8945850.

11.Q. Wu, Y. Liu, Q. Li, S. Jin, and F. Li, “The Application of Deep Learning in Computer Vision,” 2017 Chinese Automation Congress (CAC), Jinan, 2017, pp. 6522–6527.

12.D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” International Journal of Computer Vision, vol. 13, no. 2, pp. 111–122, 1981.