IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

AI Based DIY Hand Gesture Speaking System for Impaired People

Dr. R. K. Moje[1], A. A. Kalaskar[2], N. T. Kadam[3], R. H. Naikwade[4]
Assistant Professor, Department of E&TC[1]
Students, Department of E&TC[2-4]
PDEA's College of Engineering, Pune, India-412307

Abstract: *Communication is a fundamental human need, yet individuals with speech and hearing impairments often face significant challenges in expressing themselves effectively. To address this issue, we propose an AI-based Do-It-Yourself (DIY) hand gesture speaking system that bridges the communication gap by converting hand gestures into audible speech. This low- cost assistive technology leverages computer vision, deep learning, and speech synthesis to enable real-time, gesture-to- speech translation. At the heart of the system lies a camera module that continuously captures hand movements, which are then processed using image preprocessing techniques such as grayscale conversion, background subtraction, and edge detection. The refined images are fed into a convolutional neural network (CNN) trained to recognize specific gestures based on their features—such as hand shape, orientation, and finger position.*

Each recognized gesture corresponds to a specific word or phrase, which is then converted to speech using a text-to- speech (TTS) engine. The DIY nature of the system ensures it remains affordable and customizable. Components such as Raspberry Pi, Arduino boards, and standard USB cameras are used to keep the cost low and allow for user modification. The system also includes a graphical interface that displays both the detected gesture and its corresponding speech output, offering a user-friendly experience.

Keywords: Communication

I. INTRODUCTION

Sign language is a visual means of communication that uses hand gestures, facial expressions, and body movements to convey meaning. Sign language is the primary means of communication for millions of deaf people around the world, making it essential to develop technology that facilitates effective interaction between hearing and non-hearing individuals. This paper presents an innovative approach to sign language recognition by leveraging the Media Pipe Hands model and web technologies. A hand gesture recognition system provides a natural, innovative and modern way of nonverbal communication. Our system detects various hand-knuckle coordinates using Media pipe Hands Model and then passes the coordinates to KNN Classifier to classify the various hand signs into respective English language alphabets in real-time using a webcam. Media pipe pre-trained model eliminates skin-tone related biases.

II. LITRATURE REVIEW

Abdullah Mujahid, Mazhar Javed Awan, Awais Yasin, Mazin Abed Mohammed, Robertas Damaševičius, Rytis Maskeliūnas, Karrar Hameed Abdulkareem (2021) proposed a lightweight hand gesture recognition model based on YOLOv3 and DarkNet-53 CNNs. Their system achieved high accuracy in complex environments and low-resolution images. The advantages included gesture recognition without additional preprocessing or image enhancement. However, a limitation was the requirement for high computation power.[1].Their system aimed to overcome the communication barrier between sign language users and verbal speakers. The advantages of their system included portability, user-friendly interface, and cost effectiveness. However, a limitation was that their system only translated

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

words received as input into sign l anguage.[10]. effectiveness. However, a limitation was that their system only translated words received as input into sign language.[2] Sahoo, Jaya Prakash, Allam Jaya Prakash, Paweł Pławiak, and S. Samantray (2022) focused on real-time hand gesture recognition using a fine-tuned convolutional neural network (CNN). Their system aimed to develop a user-independent interface with high recognition performance. The proposed methodology included data acquisition, pre-processing techniques such as segmentation and filtering, and recognition of hand gestures using AlexNet and VGG-16. The advantages of their approach included improved segmentation with RGB- D sensors and real-time recognition. However, a limitation was the assumption that the hand is the closest object in front of the Kinect sensor.[11]. Different studies are being conducted to assess and evaluate how the device can help people with hearing and speech disabilities communicate more effectively, as well as to determine the device's limitations in contrast to other technologies and devices pursuing a similar goal. Their only means of communicating with others is through hand motions and expressions, and they have created a prosthetic speaking mouth for deaf people. This will also assist others in comprehending people with disabilities [3]. Their system aimed to capture sign gestures through a webcam, translate them into text and convert the text into speech. The advantages of their system included a 95% accuracy and the potential for extension to other sign languages with sufficient training data. However, a limitation was the need for gloves to eliminate variations in skin complexion and the possibility of incorrect predictions with bad gesture postures.[5].

III. METHODOLOGY

Block Diagram:

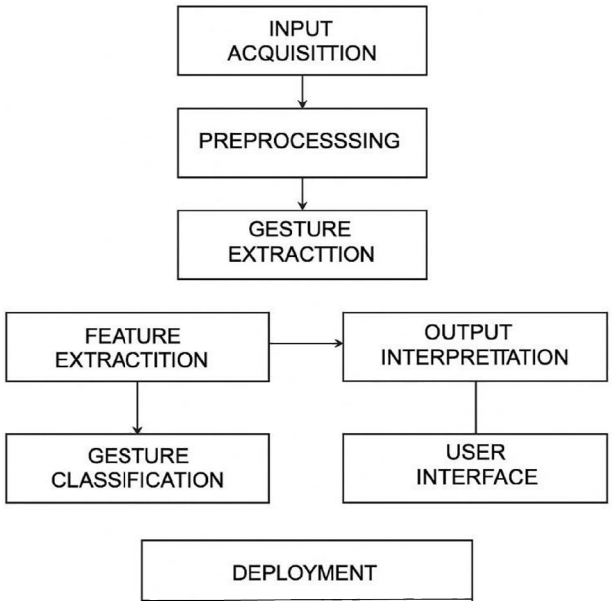


Fig. AI Based DIY Hand Gesture Speaking System

Input Acquisition:
This is the **first step** where the system captures the hand gestures using a camera or image sensor.
The system records real-time images or video frames of the user's hand movements.

Pre-processing:
In this stage, the captured images or frames undergo preprocessing to improve quality and reduce noise.
Includes operations like:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

- Grayscale conversion.
- Background removal.
- Noise filtering.
- Contour and edge detection to isolate the hand region.

Gesture Detection:
The system applies computer vision algorithms to detect the presence of a hand.
It identifies the hand shape, location, and motion.
This step prepares the image for further analysis by identifying the region of interest (ROI).

Feature Extraction:
Key features of the hand gesture are extracted, such as:

- Finger positions.
- Hand orientation.
- The distance between joints.
- These features serve as the input parameters for the machine learning model.

Gesture Classification:
The system uses a machine learning model (e.g., CNN, LSTM) to classify the hand gesture.
The model compares the extracted features with a pre-trained dataset to recognize the gesture.
Each gesture is linked to a specific label (e.g., "Hello" or "Thank You").
Output Interpretation:
The classified gesture is mapped to a corresponding output. In this case, it is associated with a text label that represents a word or phrase.
The output is prepared for speech synthesis.
User Interface:
The system includes a user interface that displays the recognized gesture and its corresponding text output.
The interface allows users to monitor the system in real time and confirm gesture accuracy.
It may also include options for customization, such as adding new gestures.

IV. OBJECTIVES

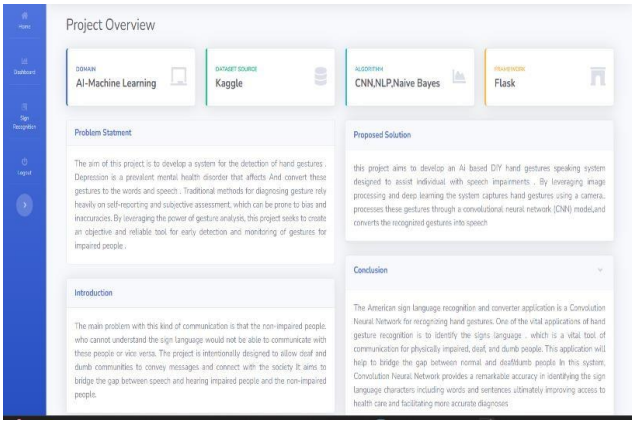
Facilitate Communication: Convert non-verbal hand gestures into spoken language, allowing speech-impaired individuals to express themselves clearly
Enhance Accessibility: Provide a low-cost, DIY solution that makes assistive communication technology affordable and customizable.
Improve Real-Time Interaction: Use AI-powered gesture recognition models to achieve fast and accurate gesture-to-speech conversion, enabling real-time communication.
Increase Independence: Empower impaired individuals by allowing them to communicate without relying on intermediaries.
Promote Portability and Usability: Use compact, wearable, and easy-to-assemble hardware (such as Arduino or Raspberry Pi) to create a portable system suitable for daily use
Enable Customization: Allow users to add or modify gestures, making the system adaptable to individual needs.
To support speech-impaired individuals in daily communication: Ultimately, the goal is to assist people who cannot speak by giving them a tool that promotes independence, enhances interaction with others, and improves their quality of life.

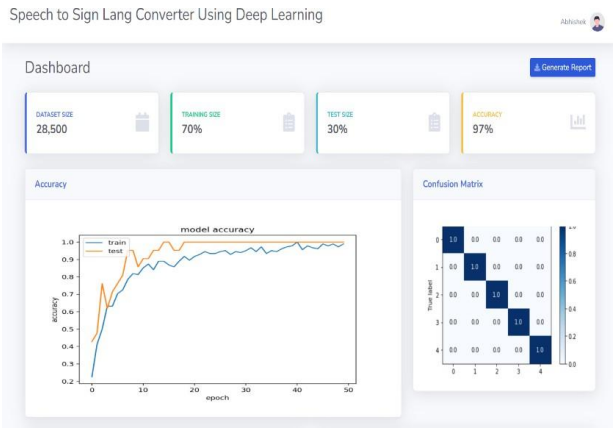
IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

Sub-Objectives

- To collect and pre-process gesture data for training the AI model.
- To analyse different hand gesture recognition techniques and identify the most efficient one for real-time systems.
- To ensure the system supports **multiple languages or dialects**, if possible.
- To test the prototype with real users and gather feedback for improvement.
- To minimize response time between gesture input and audio output.

V. PROJECT MODEL





DOI: 10.48175/IJARSCT-25913

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

VI. WORKING

The AI-based DIY hand gesture speaking system functions by capturing hand movements, interpreting them through machine learning, and converting them into audible speech. The process begins with input acquisition, where a camera captures real-time images or video frames of the user's hand gestures. These frames serve as the raw data for further processing.

Once the images are captured, they undergo pre-processing to enhance quality and remove noise. Pre-processing steps include grayscale conversion, background removal, and edge detection. These operations help isolate the hand region and highlight its contours, making it easier for the system to recognize gestures accurately.

Next, the system detects the hand gesture using computer vision techniques. It identifies the hand's position and shape, and then extracts key features such as finger orientation, palm shape, and joint distances. These features are crucial for distinguishing between different gestures.

The extracted features are then fed into a machine learning model, typically a Convolutional Neural Network (CNN), which classifies the gesture. The model is trained on a dataset of hand gestures, enabling it to recognize and label the current gesture accurately. For dynamic gestures or sequences, recurrent models like Long Short-Term Memory (LSTM) networks may be used to capture temporal patterns.

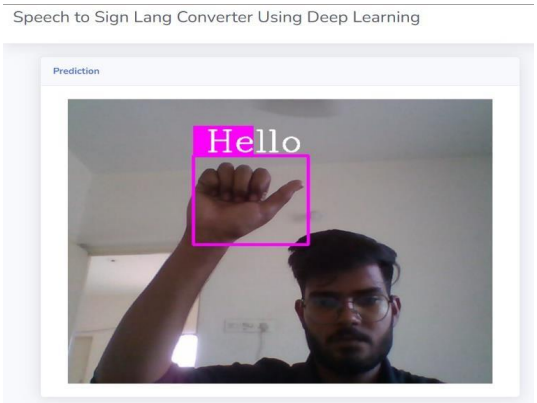
Once the gesture is recognized, the system maps it to a corresponding text label. This label represents a word or phrase associated with the gesture. The text is then converted into speech using a Text-to-Speech (TTS) module. The TTS engine synthesizes the text into natural-sounding speech, allowing the user to communicate through their hand movements.

Finally, the system displays the recognized gesture and the corresponding speech output through a user interface. This interface allows the user to monitor the accuracy of the system and make adjustments if necessary. The system can be refined and improved through iterative testing, ensuring consistent and reliable performance. Upon successful testing, it is deployed for real-world use, offering an effective communication tool for individuals with speech or hearing impairments

VII. APPLICATIONS

- Healthcare and Rehabilitation
- Educational Tools
- Security and Authentication
- Emergency Communication
- Public Services
- Industrial Automation

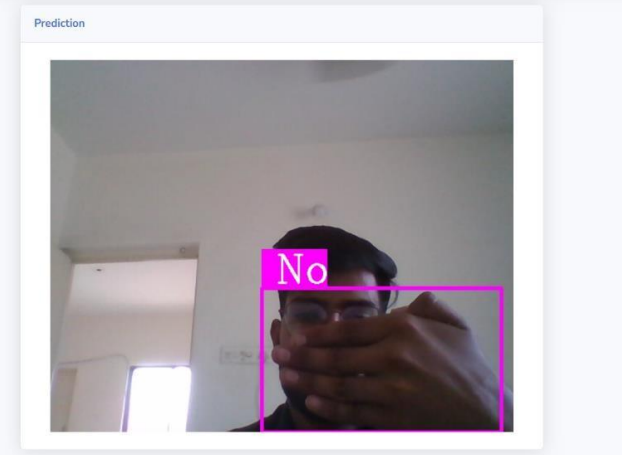
VIII. RESULTS



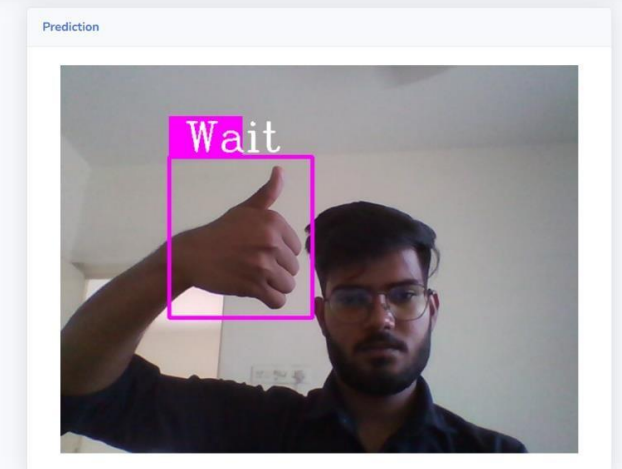
Result for: Hello

DOI: 10.48175/IJARSCT-25913

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67



Result for: No



Result for: Wait

IX. CONCLUSION & FUTURE SCOPE

The AI-based DIY hand gesture speaking system offers an innovative and practical solution for individuals with speech or hearing impairments, enabling them to communicate effectively through hand movements. By leveraging computer vision, machine learning, and text-to-speech (TTS) technologies, the system accurately recognizes hand gestures and translates them into audible speech.

The DIY nature of the system makes it cost-effective, accessible, and customizable. Users can easily build and modify the system using readily available hardware and open- source software, allowing for flexibility in gesture recognition and speech mapping. The iterative testing and refinement process ensures improved accuracy and reliability, making the system adaptable for real-world applications.

Overall, this solution enhances inclusivity by providing a voice to those with communication challenges. It demonstrates the potential of AI-powered assistive technologies in improving the quality of life, fostering independence, and promoting seamless interaction in everyday scenarios. The focus of our study was on building a real-time system that could identify hand motions to improve communication for individuals with hearing disabilities and the automatic recognition of sign gestures by our system using advanced algorithms combined with convolutional neural network technology allows for more intuitive and natural communication between individuals. Recognizing

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

commonalities among human hands' shape is essential for the system's ability to identify different types of gestures. Using our system as a tool to improve accessibility and streamline communication is an effective way to save time for all parties involved, and our target moving forward is to improve upon the system's precision and speed as well as make it more accessible for users. Developing more advanced applications integrated with modern tech such as virtual and augmented reality is our aim in enhancing user experience. Furthermore, continued research and development of the hand gesture recognition system has the potential to change communication method's

REFERENCES

[1]. Abdullah Mujahid, Mazhar Javed Awan, Awais Yasin, Mazin Abed Mohammed, Robertas Damaševi˘cius, Rytis Maskeliunas and Karrar Hameed Abdulkareem, Real-Time Hand Gesture Recognition Based on Deep Learning YOLOv3 Model, Appl. Sci. 2021.

[2]. Simei G. Wysoski, Marcus V. Lamar, Susumu Kuroyanagi Akira Iwata, (2002). "A RotationInvariant Approach On Static- Gesture Recognition Using Boundary Histograms And Neural International.

[3]. Dix A, Finlay J, Abowd GD and Beale R (2003), Human-Computer Interaction. 3rd ed. Prentice Hall.

[4]. S. Mitra, and T. Acharya. (2007). "Gesture Recognition: A Survey" IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and reviews, vol. 37 (3), pp. 311- 324, doi:10.1109/TSMCC.2007.893280

[5]. G. R. S. Murthy, R. S. Jadon. (2009). "A Review of Vision Based Hand Gestures Recognition,"International Journal of Information Technology and Knowledge Management, vol. 2(2), pp. 405-410.

[6]. Novogrodsky, R., Fish, S.,& Hoffmeister, R. (2014). The Acquisition of Synonyms in American Sign Language (ASL): Toward a Further Understanding of the Components of ASL Vocabulary Knowledge. Sign Language Studies, 14(2), 225–249.

[7]. Sign Language Translation Using Deep Convolutional Neural Networks. (2020). KSII Transactions on Internet and Information Systems, 14(2).

[8]. Rudra Kumar, M., Rashmi Pathak, and Vinit Kumar Gunjan. "Diagnosis and Medicine Prediction for COVID-19 Using Machine Learning Approach."Computational Intelligence in Machine Learning: Select Proceedings of ICCIML 2021. Singapore: Springer Nature Singapore, 2022. 123-133

[9]. Alsaadi, Z., Alshamani, E., Alrehaili, M., Alrashdi, A. A. D., Albelwi, S., & Elfaki, A. O. (2022). A Real Time Arabic Sign Language Alphabets (ArSLA) Recognition Model Using Deep Learning Architecture. Computers, 11(5),78

[10]. Tanmay Petkar, Tanay Patil, Ashwini Wadhankar Vaishnavi Chandore, Vaishnavi Umate, Dhanshri Hingnekar, "Real Time Sign Language Recognition System for Hearing and Speech Impaired People", Ijraset Journal For Research in Applied Science and Engineering Technology (Volume 10 - Issue 4, April 2022).

[11]. Sahoo, Jaya Prakash, Allam Jaya Prakash, Paweł Pławiak, and Saunak Samantray. 2022. "Real-Time Hand Gesture Recognition Using Fine-Tuned Convolutional Neural Network" Sensors 22, no. 3: 706. https://doi.org/10.3390/s22030706

[12]. Pallavi Nagarkar, Pravar Chaturvedi, Bhumika Neole Hand Gesture Speaking Unit for Mute People," International Journal of Computer Sciences and Engineering, vol. 7, no. 4, pp. 532-35, 2019.

[13]. Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur, Dr Dayananda P, 2020, Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCAIT – 2020 (Volume 8 – Issue 15)