Real-Time Gesture Recognition System for Speech Impaired People

Mr. Sahil A. Mujawar[1], Dr. Sangram T. Patil[2], Dr. Jaydeep B. Patil[3], Mr. Rushikesh R. Powar[4]

[1] Student Department of CSE(DS), D Y Patil agriculture and technical University Talsande Kolhapur

[2] Dean, School of Engineering & Technology, D Y Patil agriculture and technical University Talsande Kolhapur

[3] Associate Professor & HoD CSE, D Y Patil agriculture and technical University Talsande Kolhapur

Abstract – In today's modern society, effective communication plays a crucial role in human interaction. However, individuals who experience speech impairments encounter considerable difficulties in expressing themselves and establishing connections with others. In response to this challenge, we introduce an innovative Real-Time Sign Gesture Recognition System tailored for individuals with speech impairments. This Proposed work is designed to empower speech-impaired individuals by facilitating seamless communication with their environment in real-time.

The Sign recognition system captures the sign gestures either using camera or sensors and interpreted into meaningful sentences. Subsequently, these interpreted sentences are converted into audible speech, enabling speech-impaired individuals to participate in fluid and meaningful communication. This system useful to general people to understand the speech impaired peoples what they want to say and convey.

This paper aims to explore the sign-languages and developing the real-time sign language recognition system.

The background section focuses on evolution of sign language and a review of researchers in the development process. It also includes the new approach of recognizing the signs and converting them into the multilingual speech and operating the IOT devices with specific commands. In the next section it includes proposed architecture and new approach of converting sign language and controlling the IOT devices.

Keywords – Computer-Vision, Convolutional Neural Network, Gesture Recognition, TTS, IOT

I.INTRODUCTION

In our communities, there exists a substantial portion of individuals who depend on sign language for communication due to their inability to speak verbally. Sign language, a sophisticated method of gestural communication, utilizes hand movements, facial expressions, and body postures to convey messages. However, this presents challenges in interacting with those who are unfamiliar with sign language.

To tackle this communication barrier, we propose the Sign Language to Speech Conversion System. The primary objective of this system is to interpret finger spelling [2], which is a fundamental aspect of sign language, and converted into audible speech. This would facilitate seamless communication between speech impaired individuals and the general population.

In developing a vision-driven human hand gesture recognition framework, the complexity arises from the vast number of variables within the image space. It is imperative to extract the crucial features of the image for effective implementation. Typically, creating a robust hand gesture detection system necessitates an extensive training dataset and modeling different gestures. However, we propose a hand gesture recognition system based on CNN technology to enhance recognition accuracies with minimal effort in modeling different gestures [9].

II.BACKGROUND

The development of sign languages is indeed a natural process that has occurred within Deaf communities over generations. These languages possess all the essential linguistic components found in spoken languages, including grammar, syntax, vocabulary, and cultural nuances. Much like spoken languages, sign languages have evolved organically within their respective communities. Much like spoken languages, sign languages have evolved organically within their respective communities.

However, one notable figure in the history of sign language is Abbe Charles-Michel de l'Épée, a French educator who founded the first public school for the deaf in Paris in the late 18th century. He is often credited with developing a systematic approach to teaching sign language and with popularizing its use in education for the deaf.

One notable author whose work focuses on real-time sign language recognition systems is Dr. Muhammad Asif. Dr. Asif's research has contributed significantly to the development of real-time sign language recognition systems. In his Dr. Asif provides an extensive review of various techniques, methodologies, and challenges in real-time sign language recognition [6].

III.PROPOSED SYSTEM

A) Problem Statement:

Speech impairment presents significant obstacles to effective communication and social interaction for individuals reliant on alternative modes of expression, such as sign language. While sign language facilitates communication within the deaf and hard-of-hearing community, it creates barriers when interacting with those unfamiliar with it. Traditional solutions, like written communication or interpreter services, often prove cumbersome, time-consuming, and may not be universally available [2]. The absence of real-time communication tools tailored to speech-impaired individuals exacerbates these challenges, limiting their ability to express themselves fluidly and engage in spontaneous conversations. Existing gesture recognition systems may lack the accuracy, speed, or adaptability required for natural and intuitive communication in real- world scenarios. Furthermore, the absence of integrated features such as language translation and IoT device control further hampers the functionality and usability of these systems [1].

Hence, there exists an urgent need for an advanced real-time gesture recognition system specifically designed for speech-impaired individuals [4]. Such a system should accurately interpret hand gestures, construct coherent sentences in natural language, and convert them into audible speech in real-time. Additionally, it should support multiple languages and enable users to seamlessly operate IoT devices through gesture-based commands.

B) Proposed System:

The proposed real-time gesture recognition system for individuals with speech impairments utilizes a CNN model to capture images via a camera interface. This specialized CNN model undergoes training to identify characters formed by hand gestures, allowing the system to interpret gestures as distinct elements of a sentence. Through organizing these identified characters, the system constructs coherent sentences, thereby facilitating effective communication.

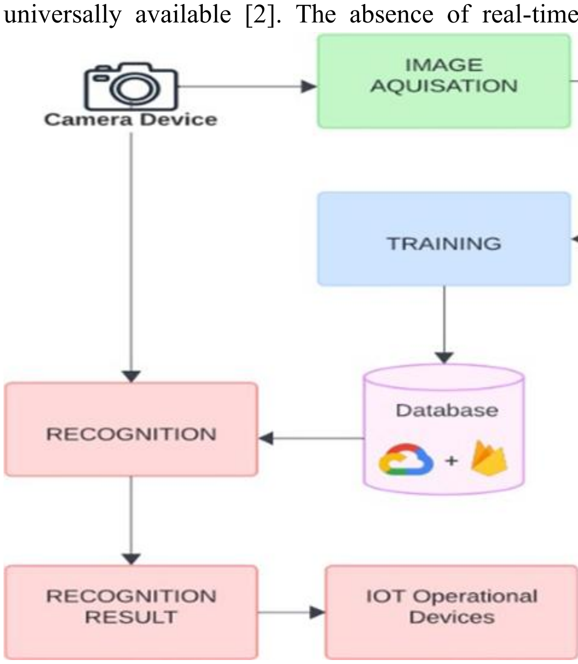

Fig1. System Architecture

The essential components of this system include:

1. Training Dataset Creation
2. Finger Spelling Recognition Module
3. Text-To-Speech Conversion Module
4. Multilingual Transcription Module
5. Operating the IoT Devices

Subsequent to gesture recognition and sentence construction, the system incorporates a text-to-speech conversion module. This module provides users with the option to select their preferred language for speech synthesis, enabling seamless translation of the constructed sentences into spoken language. This functionality significantly enhances accessibility and inclusivity by accommodating users' language preferences [7].

Furthermore, the system expands its capabilities to include the operation of Internet of Things (IoT) devices using specific commands integrated into the constructed sentences. By incorporating commands for IoT device control within recognized gestures and sentences, individuals with speech impairments acquire the ability to interact with and control their environment. This enhancement promotes their autonomy and independence in daily activities [3].

Dataset Creation

For developing such a big system, we require a large dataset to work with. The creation of a high-quality dataset is paramount to the success of the gesture recognition system.

The dataset should cover a wide range of hand gestures commonly utilized in sign language, ensuring that the Convolutional Neural Network (CNN) model is trained on a comprehensive array of gestures to achieve robust recognition performance.

The dataset creation process involves several key steps:

1. Data Acquisition: Hand gesture images are captured utilizing a high-resolution camera or depth sensor. Care is taken to capture images under various lighting conditions, angles, and backgrounds to replicate real-world variability [8].

(a) (b) (c)

Fig 2. Dataset Creation (Gesture Recognition)

2. Annotation: Each captured image is annotated with the corresponding label or gesture class. This annotation step is crucial for supervised learning, as it furnishes ground truth labels for training the CNN model.

3. Data Augmentation: To enhance the diversity and robustness of the dataset, data augmentation techniques such as rotation, scaling, translation, and noise addition may be employed on the captured images. This helps mitigate overfitting and enhances the CNN model's generalization capability.

4. Dataset Split: The dataset is partitioned into training, validation, and test sets to facilitate model training, hyperparameter tuning, and performance evaluation, respectively. Attention is paid to maintaining a balanced distribution of gesture classes across the different sets.

IV.CNN (CONVOLUTIONAL NEURAL NETWORK)

Convolutional Neural Networks (CNNs) represent a class of neural networks particularly effective in addressing computer vision challenges. They draw inspiration from the human visual system's perception, specifically the processes that occur in the visual cortex. CNNs employ filters or kernels to traverse the pixel values of an image, assigning appropriate weights to facilitate the detection of specific features. Comprising various layers such as convolutional, max pooling, flatten, dense, dropout, and fully connected neural network layers, CNNs form a potent tool for feature identification in images. Distinct from traditional Neural Networks, CNN layers organize neurons in three dimensions: width, height, and depth. Neurons within a layer establish connections limited to a small region (window size) of the preceding layer, as opposed to a fully-connected arrangement encompassing all neurons [10].

In CNN architecture, initial layers detect low-level features, progressively transitioning to higher-level features of increasing complexity. This hierarchical feature extraction enables CNNs to effectively interpret image content.

In the convolution layer, we employ a compact window dimension, generally 5x5, which spans through the depth of the input array. Inside this layer, there are trainable kernels of identical window size. During each iteration, we shift the window by a step size, usually 1, and calculate the dot product between the filter entries and the input data at a specific location. This repeated procedure produces a 2D activation map, representing the response of the matrix across different spatial locations. Fundamentally, the network develops filters that respond to certain visual patterns, like edges in different directions or regions with distinct colors. [6].

V.RESULTS

The fig. 3(a), 3(b) and 3(c) shows sample outputs from the system, how it captures gesture and convert into characters and finally forms the relevant sentence. Now the message is conveyed through multi-lingual voice to others.

Fig. (a)

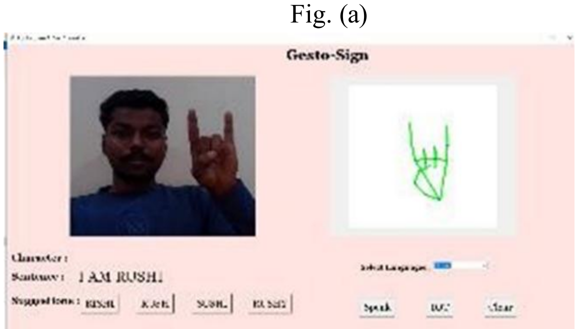

Fig. (b)

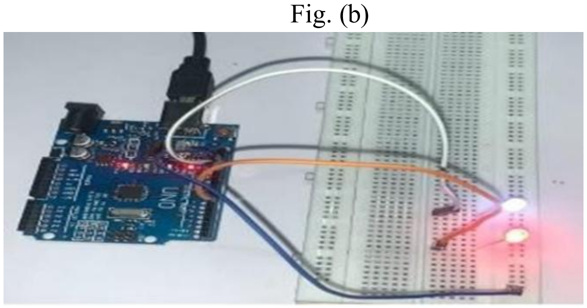

Fig. (c)

Fig 3. Output of Working Model

Then to check the accuracy of the model, we carried out some experiments and the details are given in table 1. The experiments are carried out for character recognition based on parameters like attempts performed, no. of times character recognized identified correct, no. of times failed to recognize character in a given amount of time and accuracy level. The Fig.4 represents the character recognition accuracy for sample characters. The experiment shows that the proposed system recognizes the most of characters with 100% accuracy and few characters like P, Q gives less accuracy as compared to other characters and the average accuracy is 95.21%.

Table 1. Character Recognition Accuracy

Sr. No.	Character	Attempt	Pass	Fail	Accuracy in %
1	A	10	10	0	100
2	B	10	10	0	100
3	C	10	9	1	90
4	D	10	10	0	100
5	E	10	10	0	100
6	F	10	10	0	100
7	G	10	9	1	90
8	H	10	9	1	90
9	I	10	8	2	80
10	J	10	8	2	80
11	K	10	10	0	100
12	L	10	10	0	100
13	M	10	10	0	100
14	N	10	10	0	100
15	O	10	10	0	100
16	P	10	7	3	70
17	Q	10	6	4	60
18	R	10	10	0	100
19	S	10	10	0	100
20	T	10	10	0	100
21	U	10	10	0	100
22	V	10	10	0	100
23	W	10	10	0	100
24	X	10	10	0	100
25	Y	10	10	0	100
26	Z	10	8	2	80
			Total Accuracy (%)		95.21

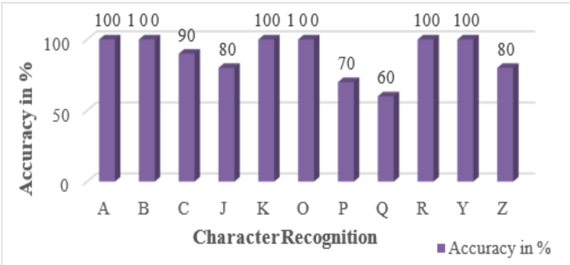

Fig. 4. Character Recognition Accuracy

VI.FUTURE SCOPE

The future prospects for the proposed Real-Time Gesture Recognition System tailored for individuals with speech impairments are extensive and encompass a range of avenues for advancement. A primary focus area involves enhancing the accuracy of gesture recognition, achievable through the refinement of Convolutional Neural Network (CNN) model architecture and the exploration of sophisticated computer vision algorithms. Moreover, upcoming iterations could integrate multimodal communication features by incorporating facial expressions or body movements, thereby enriching the communicative experience. Introducing personalized functionalities to enable users to tailor the system according to their preferences is another potential enhancement, alongside mechanisms for continuous learning and adaptation to refine the system over time based on user input. Integration with wearable devices emerges as a promising direction, facilitating hands- free operation and improving accessibility for users with limited mobility. Expanding language support [10], integrating with smart home ecosystems and IoT devices, and implementing real-time feedback mechanisms are additional pathways to enhance the system's efficiency and efficacy. Overall, sustained innovation and interdisciplinary collaboration will propel the ongoing development of the Real-Time Gesture Recognition System, ensuring its ability to meet the evolving requirements of individuals with speech impairments while enhancing their communication experiences.

VII.CONCLUSION

In summary, the Real-Time Gesture Recognition system can be used by common people to understand and learn sign language. This will help them to interact and have a communication with speech impaired people introduced, in this study signifies a significant leap forward in assistive technology tailored for individuals with speech impairments. By harnessing CNN technology, the system enables instantaneous identification of sign language gestures, thereby facilitating seamless and intuitive communication. Furthermore, the incorporation of multilingual text-to-speech conversion and IoT device management amplifies the system's capabilities, providing users with efficient tools for interaction and control. Future research endeavours may involve refining the CNN model to enhance its precision and efficiency, broadening language support, and further integrating with IoT devices to augment the system's accessibility and usability. In essence, this study contributes to the advancement of innovative solutions that address the distinctive communication obstacles encountered by speech-impaired individuals, thereby fostering inclusivity and empowerment within society.

REFERENCE

[1] Deshpande, Aditi, Ansh Shriwas, Vaishnavi Deshmukh, and Shubhangi Kale. "Sign Language Recognition System using CNN." In 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), pp. 906-911. IEEE, 2023.
[2] Ziaie, Pujan, Thomas Müller, Mary Ellen Foster, and Alois Knoll. "A naïve Bayes classifier with distance weighting for hand-gesture recognition." In Advances in Computer Science and Engineering: 13th International CSI Computer Conference, CSICC 2008 Kish Island, Iran, March 9-11, 2008 Revised Selected Papers, pp. 308-315. Springer Berlin Heidelberg, 2009.
[3] Mohamed, Tamer, Amer Ibrahim, Tauqeer Faiz, Waseem Alhasan, Ayesha Atta, Vansh Mago, Muhammad Ahzam Ejaz, and Salman Munir. "Intelligent Hand Gesture Recognition System Empowered With CNN." In 2022 International Conference on Cyber Resilience (ICCR), pp. 1-8. IEEE, 2022.
[4] Kausar, Sumaira, and M. Younus Javed. "A survey on sign language recognition." In 2011 Frontiers of Information Technology, pp. 95-98. IEEE, 2011.
[5] Bantupalli, Kshitij, and Ying Xie. "American

sign language recognition using deep learning and computer vision." In 2018 IEEE International Conference on Big Data (Big Data), pp. 4896-4899. IEEE, 2018.

6] Abiodun, Oludare Isaac, Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Abubakar Malah Umar, Okafor Uchenwa Linus, Humaira Arshad, Abdullahi Aminu Kazaure, Usman Gana, and Muhammad Ubale Kiru. "Comprehensive review of artificial neural network applications to pattern recognition." IEEE access 7 (2019): 158820- 158846.

[7] Kania, Marta, Ewa Korzeniewska, Rafał Zawiślak, Alyona Nikitina, and Andrzej Krawczyk. "Wearable Solutions for the Sign Language." In 2022 IEEE 4th International Conference on Modern Electrical and Energy System (MEES), pp. 1-4. IEEE, 2022.

[8] Halder, Arpita, and Akshit Tayade. "Real-time vernacular sign language recognition using mediapipe and machine learning." Journal homepage: www. ijrpr. com ISSN 2582 (2021): 7421.v

[9] Taskiran, Murat, Mehmet Killioglu, and Nihan Kahraman. "A real-time system for recognition of American sign language by using deep learning." In 2018 41st international conference on telecommunications and signal processing (TSP), pp. 1-5. IEEE, 2018.

[10] Pigou, Lionel, Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen. "Sign language recognition using convolutional neural networks." In Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13, pp. 572-578. Springer International Publishing, 2015.