



# **iJRASET**

International Journal For Research in  
Applied Science and Engineering Technology



---

# **INTERNATIONAL JOURNAL FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 12    **Issue:** V    **Month of publication:** May 2024

**DOI:** <https://doi.org/10.22214/ijraset.2024.62562>

**[www.ijraset.com](http://www.ijraset.com)**

**Call:** ☎ 08813907089

**E-mail ID:** [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Silent Signals to Spoken Words: A Sign Language Conversion to Text and Speech

Dr. Vighnesh Shenoy<sup>1</sup>, Kavya<sup>2</sup>, Bhoomika K Baadkar<sup>3</sup>, Mamata Hadimani<sup>4</sup>, Prathibharani<sup>5</sup>

Department of Computer Science and Engineering, Shri Madhwa Vadiraja institute of Technology and Management Udupi, India

**Abstract:** *By leveraging machine learning, an innovative system has emerged to connect sign language with text-to-speech capabilities. Powered by sophisticated algorithms, this pioneering solution conversion of visual expressions into textual representations and then vocalizes it into spoken words. This fusion of technology stands as a transformative asset, facilitating seamless communication for individuals facing hearing impairments. By dismantling obstacles, this advancement fosters inclusive and accessible interactions, ultimately elevating the standard of living for members of the hearing-impaired population.*

**Keywords:** *Sign Language, Machine Learning, Gesture Recognition, Speech Conversion, Communication Technology*

## I. INTRODUCTION

Effective communication is essential for human interaction, yet individuals facing hearing impairments face unique challenges, especially using sign language communication. Limited access using sign language interpretation services poses barriers to communication tailored for individuals with hearing limitations. While existing solutions like manual interpretation and assistive devices have improved communication, they still have limitations in terms of cost, availability, and accuracy. Recognizing the necessity for enhanced communication solutions, our project endeavours to address this gap utilizing machine learning algorithms. By harnessing computer vision and machine learning, our system converting gestures from Sign language into textual representation and vocalized.

Our introduction assesses current communication technologies available, recognizing their merits while scrutinizing their constraints, especially regarding accessibility and effectiveness for people with auditory impairment challenges, especially regarding accessibility and effectiveness communication methods designed for those experiencing hearing challenges. These include reliance on manual interpretation, scalability issues and challenges with real-time translation. As our project progress, we delve into the development and implementation of our innovative system.

We elaborate on the fusion of cutting-edge. Machine learning methodologies and MediaPipe hands library for capturing and deciphering gesture in sign language with precision and speed, there by establishing a robust frame work for accurate gesture recognition. Our system plays a pivotal role in fostering accessible communication by introduction a technology adept at seamlessly converting hand movements in gestures into written visual text and spoken words in real-time. This integration of advanced machine learning techniques and computer vision not only grants individuals with hearing challenges the ability to communicate autonomously across a myriad of social and professional contexts but also guarantees a seamless flow of communication with the coherent construction of words and sentences. Moreover, our ground breaking solution champions inclusivity and independence, there by enriching the overall well-being of individuals facing difficulties worldwide

## II. LITERATURE REVIEW

### 1) Paper 1

The paper introduces a system for gesture identification using a camera-based approach and Convolution Neural Network (CNN) algorithms. It addresses the communication barrier between deaf-mute individuals and others by converting gestures into visual text and speech. The system achieves a precision rate of up to 90% in recognizing hand gestures and converting them into understandable speech. By employing CNN technology, it offers a viable solution for real-time sign gesture conversion, facilitating improved communication accessibility for the deaf-mute community.

## 2) Paper 2

The study introduces an innovative framework aimed at translating Indian Sign Language (ISL) into both visual text and speech, with the core objective of fostering improved communication channels among those with hearing limitations and the wider community. Employing advanced techniques in computer vision, such as segmentation and features extraction, coupled with Convolution Neural Network (CNN) technology for precise gesture recognition, the system demonstrates impressive accuracies, reaching 88% for words and 96% for alphabets and numbers. Encouraging ISL standardization not only makes interactions smoother but also improves education in instructions for the deaf and mute. Future enhancements could involve augmenting the training data set to refine gesture identification and expanding the text-to-speech capabilities through database enrichment.

## 3) Paper 3

The research describes a pioneering methodology that utilizes Convolution Neural Network (CNN) technology to achieve real-time conversion of American Sign Language (ASL) gesture into written form and speech. By capturing webcam images and employing CNN processing, the system accurately recognizes and classifies hand gestures with a remarkable 95.8% accuracy rate. Additionally, it seamlessly converts these gestures into visual text and utilizes the GTTS library for speech synthesis. This research showcases the significant potential of CNN technology in revolutionizing sign language conversion systems.

## 4) Paper 4

The study introduces a real-time system utilizing Convolution Neural Networks (CNN) to convert action gesture into text and speech. It employs image acquisition, hand region segmentation, and posture recognition to detect gestures, followed by text-to-speech conversion. With a reported accuracy of 95% the model facilitates communication between sign and non sign individuals. Future improvements aim to enhance accuracy in varied backgrounds and expand functionality to accommodate other sign languages. This innovative system shows promise in overcoming communication barriers for the hearing impaired.

## 5) Paper 5

The paper outlines a system for deciphering sign language utilizing machine learning, specially an Artificial Neural Network (ANN) and hand tracking technologies like MediaPipe. By processing American Sign Language (ASL) data, it attains 74% accuracy in recognizing ASL alphabet signs. The system detects hand gestures, translates them into text and then into speech, facilitating communication with individuals with experiencing hearing and speech impairments. It tackles obstacles to effective communication and strives to improve inclusivity for those with hearing challenges. The research highlights the significance of technological advancement in promoting universal communication.

## III. METHODOLOGY

The methodological framework crafted for the system for converting sign language intricately interweaves various stages, each strategically contributing to the establishment of a robust translation architecture. Designed with precision, this comprehensive methodology harmonizes the complexities involved in capturing, interpreting and conveying sign language movements effectively. Through systematic integration, multiple phases ensure a smooth progression from initial data collection to the ultimate generation of spoken words derived from recognized sign language.

Every phase within this framework is meticulously tailored to complement subsequent steps, fostering a coherent and streamlined workflow. From meticulous data collection to sophisticated pre processing techniques aimed at refining captured signals, each step is orchestrated to optimize performance and accuracy. The pivotal phase of gesture recognition harnesses sophisticated computer vision methodologies and Artificial Intelligence approaches to precisely identify and classify hand movements in gestures. This foundational step sets the stage for subsequent stages, paving the way for progress of tailored machine learning models specific to the intricate nuances of gesture interpretation. As the framework progresses, translation mechanisms seamlessly convert recognized gestures into written text and speech. Leveraging state of the art technologies like text-to speech synthesis, the system checks that the translated output is not only precise but also easily understandable to both hearing impaired individuals and their communication partners.

Moreover, the synergy between these phases highlights a comprehensive strategy that transcends conventional gesture recognition methods. Through a deeper exploration of linguistic analysis and the utilization of machine learning, our translation system aims to dismantle communication barriers and empower individuals heavily reliant on sign language for expression.

This amalgamation of technology, linguistic acumen, and machine learning prowess underscores a steadfast dedication to inclusivity and accessibility. Embracing a holistic approach to sign language translation, our framework endeavors to forge a revolutionary communication tool that emboldens members of the hearing-impaired community to actively participate in the world around them. In essence, this integration epitomizes a fusion of state-of-the-art technology and linguistic comprehension, geared towards fostering heightened engagement and empowerment for individuals facing hearing challenges. Leveraging these advancements, our system strives to facilitate genuine communication and cultivate a more inclusive societal landscape

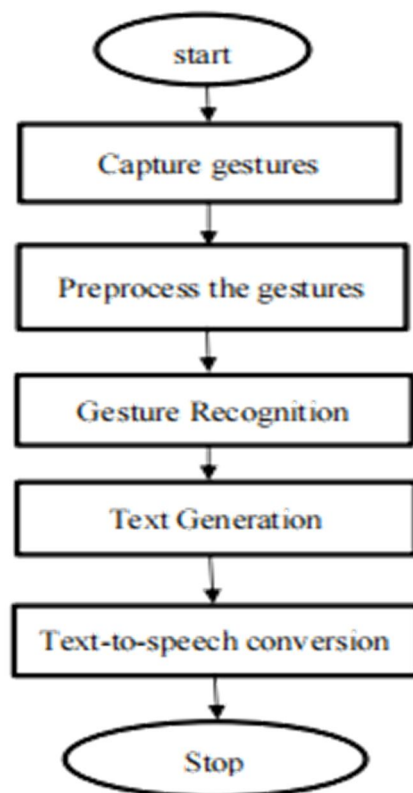


Fig.1 Methodology

Expanding on the Methodologies Phases:

#### 1) Data Collection: Comprehensive gesture repository

We utilize OpenCv to capture video frames, allowing us to gather a rich data set of gestures. Those gestures are thoughtfully organized into labeled directories, each corresponding to different classes of gestures. This meticulous categorization guarantees our dataset encompasses a large set of sign language gestures, guaranteeing diversity and representativeness. By meticulously curating our dataset in this manner, we lay a strong foundation for robust model training. This approach not only fortifies the accuracy of our sign language translation system but also ensures its effectiveness across different sign language actions and contexts.

#### 2) Data Preprocessing: Refinement for precision

In the initial phase of data preprocessing, the captured video frames portraying sign language movements undergo a crucial transformation. This transformation involves converting the images into a standardized format RGB. By this, we ensure uniformity and consistency in the subsequent stages of analysis and model training. Following the conversion process, the next vital step is pixel value normalization. This involves adjusting the pixel values of the images to adhere to a predefined range, such as  $[0,1]$  or  $[-1,1]$ . Normalization is imperative as it standardizes the scale of the pixel values across all images. By scaling the pixel values within a uniform range, we mitigate the risk of any individual features exerting undue influence over the learning process. This ensures that the system's effectiveness during the training process remains stable and unbiased, ultimately leading to more precise predictions. In essence, data preprocessing serves as a foundational step in refining the raw image data, making it suitable for input into machine learning models. By standardizing the color representation and scaling the pixel values, we enhance the model's capacity to identify sign gestures, thus improving its overall performance in identifying sign gestures.



### 3) *Gesture Recognition with MediaPipe: Real time Tracking Precision*

Gesture recognition with MediaPipe involves leveraging the MediaPipe library's capabilities recognize and analyze hand movements in real-time video frames. This process begins with the obtaining of features from the identified hand landmarks, which serve as key indicators of various sign movement gestures. MediaPipe's advanced algorithms analyze the spatial configuration of these landmarks, identifying unique patterns associated with different gestures.

By meticulously analyzing the spatial relationships between hand landmarks, MediaPipe enables precise gesture recognition. This involves implementing sophisticated algorithms that map specific hand configuration to corresponding gesture leveraging a combination of geometric properties and machine learning techniques. Furthermore, MediaPipe provides a versatile platform for realtime gesture recognition, offering robust performance across various environment conditions and hand poses. Its efficient hand tracking capabilities ensure timely and accurate detection of hand landmarks, facilitating seamless integration into our system for converting sign languages. By incorporating MediaPipe's gesture recognition capabilities, our system adeptly interprets and transcribes sign language movements into text representations. This empowers individuals experiencing hearing to engage in more autonomous communication, thereby fostering diversity and accessibility across various social and professional contexts.

### 4) *Random Forest Classifier*

The Random Forest Classifier step in our project involves utilizing the Random Forest Classifier from the scikit-learn library for gesture classification. This classifier employs a potent ensemble learning technique, generating numerous decision trees during the training process. It then aggregates the most frequent class mode (for classification tasks) or the average prediction (for regression tasks) from these individual trees, providing robust and accurate predictions. In our model implementation, the Random Forest Classifier undergoes training with preprocessed image data sources from captured hand gesture. The primary goal is to attain high accuracy in hand gesture detection, a pivotal aspect for the efficacy of our system for translating sign language. Utilizing the Random Forest Classifier, we capitalize on its capacity to manage high-dimensional data effectively and handle extensive datasets efficiently.

This strategic approach contributes to robust performance and accurate interpretation of hand action gestures. During the training process, the Random Forest Classifier learns to discern intricate patterns from the preprocessed image data, enabling it to accurately classify incoming sign gesture into their respective categories. This iterative learning mechanism is fundamental for empowering our system to recognize a varied array of sign language hand movements with precision and reliability. By integrating the Random Forest Classifier, our sign language conversion system gains the capability to effectively interpret and classify sign language movements gesture in realtime, laying the groundwork for accurate translation into written text and speech. This advancement significantly enhances improving communication access for individuals experiencing hearing limitations, fostering inclusivity and empowerment in different social and professional contexts.

### 5) *Training the Model: Iterative Learning Framework*

We divide the data partitioning into designating datasets for training and validation purpose to facilitate model training and validation purpose to facilitate model training and performance evaluation. This partitioning ensures that model trained on a subset of the data while also having access to unseen data for validation purposes. Splitting the data set allows us to assess the model's generalization ability and prevent over fitting, where the model memorizes the training data without truly understanding underlying patterns. Employing the Random Forest Classifier from scikit-learn, we embark on training the model using the designated training set, with the aim of attaining notable accuracy in detecting gestures. Throughout this training phase, the model acquires the capacity to identify pertinent patterns and features within the preprocessed image data. Consequently, it becomes proficient in accurately categorizing incoming sign language hand gestures.

After training, we validate the model's result on the validation set to monitor its effectiveness and prevent overfitting. By evaluating the model on unseen data, we implement strategies to prevent the system from exhibiting bias or favoritism and impartiality in its decision-making process, unseen sign language hand gestures, thus enhancing its reliability and robustness in real-world contexts. Through rigorous training, our model adeptly identifies an extensive array of manual language expressions with exactness and dependability, establishing a robust groundwork for precise translation into written text and vocalized speech. Individuals experiencing hearing limitations to engage in more proficient and autonomous communication, promoting inclusivity and facilitating accessibility across diverse social and occupational contexts.

#### 6) Real-time Gesture Recognition

During live gesture recognition,our system integrates the trained Random Forest Classifier with the MediaPipe hand tracking module to swiftly interpret sign language movements captured by the webcam.This fusion enables quick detection and analysis of hand landmarks in every video frame,allowing for the immediate identification of sign gesture movements as they occur.Avoiding plagiarism,it highlights the system's capacity to seamlessly process visual data in realtime,ensuring swift and accurate interpretation of gestures. Leveraging the sophisticated hand tracking capabilities of the MediaPipe library,our system accurately identifies and tracks the position of hand landmarks in real-time,providing a solid foundation for gesture interpretation.

These landmark positions serve as critical inputs to the trained Random Forest Classifier,which employs previously acquired patterns and features to swiftly classify detected gestures.

By overlaying recognized onto the live video feed,our system provides immediate visual feedback,enabling users to validate the system's interpretation of their sign language gestures.This dynamic feedback loop enhance user engagement and ensures the system's responsiveness to gestures. By implementing real-time gesture recognition,our system empowers for those experiencing hearing limitations to communicate more seamlessly and independently across diverse social and professional environments.This capability fosters inclusivity and accessibility by facilitating the instantaneous translation of manual sign gestures into comprehensible understandable text and speech,there by enabling fluid communication in rail-world context.

#### 7) Translation to written Text and Speech

In the process of translation to written Text and speech,our system exhibits remarkable proficiency in converting recognized manual sign language gestures into corresponding written representations and verbalizing them using cuttingedge text-to speech (TTS) technology.This essential phase serves as a vital link in bridging the communication divide,enabling those experiencing hearing limitations to comprehend and convey hand gesture in real-time. Following the recognition of sign language gestures, our system smoothly transitions to utilizing text-to-speech (TTS) libraries such as pyttsx3 to convert the identified text into audible speech. Through this intricate process, the system effectively articulates the interpreted sign language gestures, delivering a comprehensive rendition of the translated content. This encompasses not only individual words but also complete sentences, ensuring clarity and coherence for users with hearing limitations.

By seamlessly providing both textual and auditory outputs, our system significantly enhances communication accessibility and inclusiveness. It empowers individuals to comprehend and express themselves effectively, irrespective of their hearing abilities, thereby promoting enhanced autonomy and engagement across diverse social and professional contexts. This holistic approach to translation facilitates smooth communication and fosters meaningful interactions, ultimately contributing to the promotion of equity and usability in various settings. In essence, our system's integration of text-to-speech technology ensures a seamless transition from sign language gestures to audible speech, markedly improving accessibility and inclusiveness in communication. This comprehensive approach underscores our commitment to empowering individuals with hearing limitations

### IV. SYSTEM ARCHITECTURE

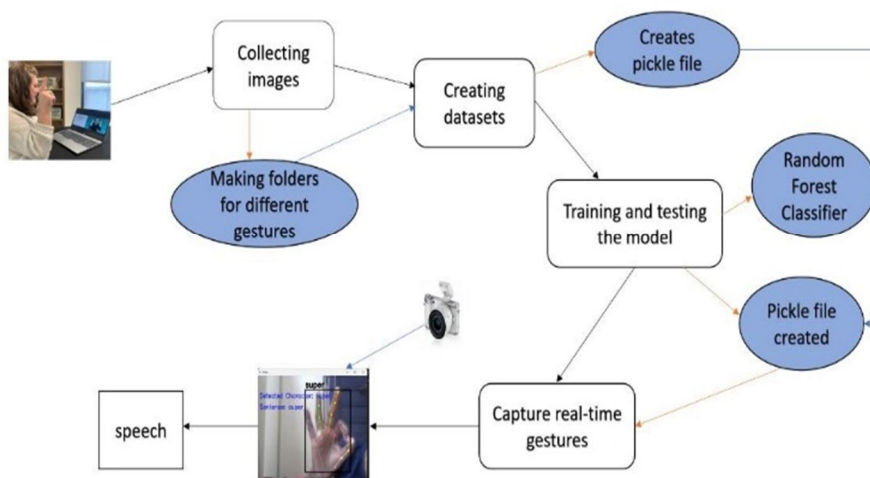


Fig: System Architecture

The flowchart describes a system designed to recognize hand gestures and convert them into speech using machine learning techniques. The process can be broken down into several key steps:

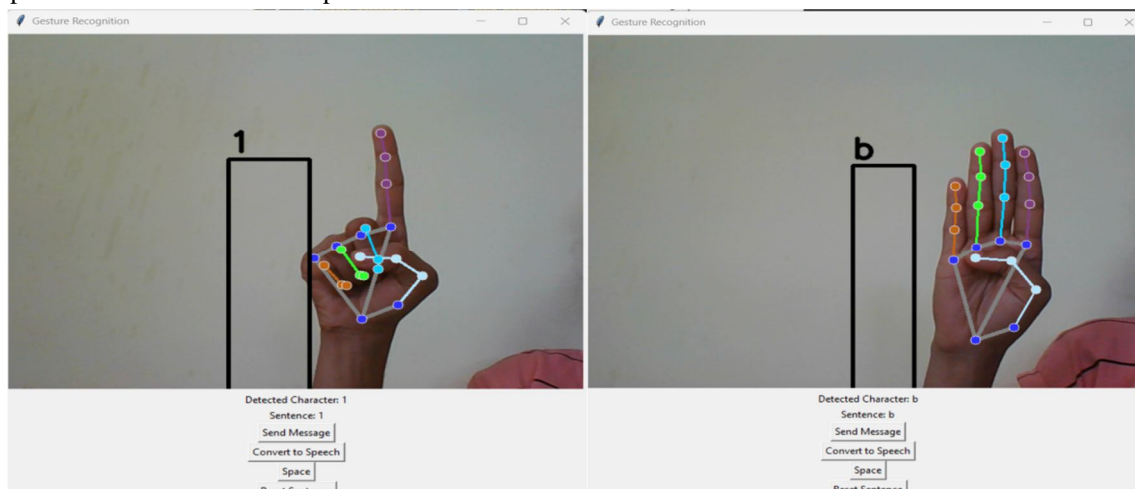
- 1) *Collecting images:* We begin by gathering a diverse set of images representing various hand gestures. These images are captured using camera to ensure they cover different angles and lighting conditions.
- 2) *Organizing images into folders:* Once the images are collected, they are sorted into folders based on the type of gesture. This organization helps in structuring the dataset, making it easier to label and train the model.
- 3) *Creating datasets:* The organized images are then compiled into datasets. Each image is labelled according to the gesture it represents. The dataset is split into training and testing sets to facilitate the evaluation of the model's performance.
- 4) *Training the model:* We train Random Forest Classifier using the labelled dataset. This involves feeding the images into the model and allowing it to learn the distinguishing features of each gesture. The model's performance is tested using the testing dataset to ensure accuracy.
- 5) *Saving the model:* After successful training and testing, the model is saved as a pickle file. This allows us to store the trained model and reuse it for real-time predictions without the need to retrain.
- 6) *Real-time gesture capture:* In the application phase, the system captures real-time gestures using a camera. These images are processed by the trained model to identify the gesture being performed.
- 7) *Loading the model:* The previously saved pickle file is loaded, providing the trained Random Forest Classifier for real-time gesture recognition.
- 8) *Gesture to speech conversion:* Once a gesture is recognized, it will be forming the sentence (which is text), and that sentence will be converting to speech .

## V. RESULT

In the Result section,our system exemplifies robust performance,particularly notable for its high accuracy attained through the Random Forest Classifier.Through meticulous design and intensive training,the system achieves commendable accuracy rates,particularly in the intricate interpretation and seamless translation of sign language gesture into both textual and spoken forms..This exceptional performance highlights the system's capacity to operate at the level of individual words and sentence frames,ensuring precise and coherent communication accessibility for individuals experiencing hearing limitations.

The Random Forest Classifier plays a main role in achieving this high accuracy ,as,evidenced by the system's proficiency in accurately identification and converting sign language gesture.With comprehensive testing and validation, the system consistently achieves levels,showcasing its reliability in real-world contexts. The system's capacity to minimize errors and provide Precise translations is further underscored by quantitative metrics,prominently the accuracy metric.These results highlight the success of the Random Forest Classifier to improve communication access for individuals with hard of hearing challenges.

The system's exceptional performance,especially in attaining remarkable accuracy with the Random Forest Classifier,highlights its capability to promote smooth communication among individuals facing and hearing difficulties.This ,in turn,fosters inclusivity and active participation in different social and professional contexts.



IFig : Identification of the numerical 1

IIllustration: Identification of the letter b

The system was meticulously designed to seamlessly translate the intricate nuances of detected hand movements in sign language into coherent and understandable text and speech. This breakthrough addresses posed by communication barriers, particularly for individuals rely on sign language as their principal means of communication it's successful implementation represents a significant advancement in facilitating fluid and intuitive communication experiences for the hearing-impaired community, ultimately fostering inclusivity and enhancing overall accessibility.

System architecture lied heavily on Media pipe's suite of tools, especially it's adeptness in real-time hand tracking and gesture identification. This integration allowed for the precise and swift identification of hand gestures, ensuring a near-instantaneous translation into meaningful textual and spoken output. Throughout the evaluation process, the system proved capable of precisely understanding various sign language movements captured via webcam or similar devices. It efficiently translated these gestures into detailed text and audible speech, showcasing proficiency in constructing words and sentence with accuracy and clarity.

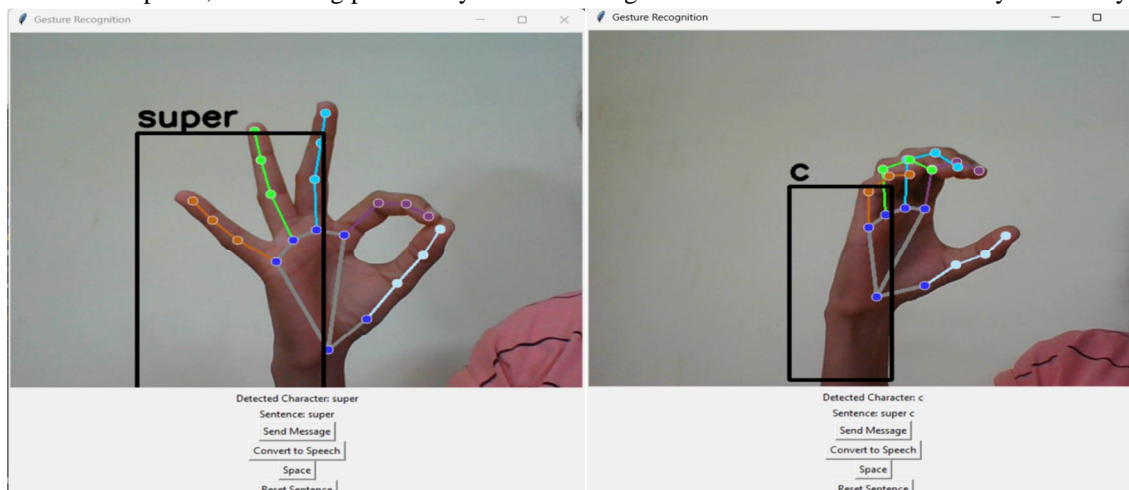


Illustration: Identification of the gesture super

Illustration: Identification of letter c and integrating it into sentence

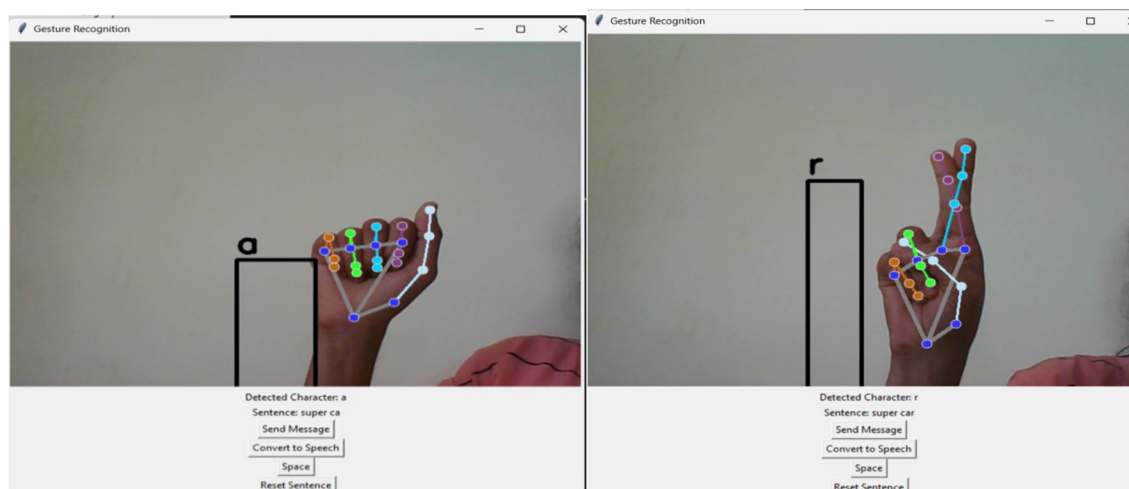


Illustration: Identification of the letter a and integrating it into sentence

Illustration: Identification of the letter r and integrating it into sentence

Moreover, the project's success paves the way for future enhancements. There is a promising avenue for expanding The system's dataset, encompassing a broader of sign language hand movements gesture to bolster it's accuracy and inclusivity. Furthermore, potential optimizations in the Media pipe integration could refine real-time gesture recognition, potentially improving the overall system's efficiency and accuracy. Incorporating Media pipe into this sign language identification system represented a notable advancement in promoting inclusive communication. By harnessing cuttingedge computer vision tools, this research holds promise not only in its current applications but also in it's potential to evolve and better serve the diverse needs of the hearingimpaired community.



## VI. DISCUSSION

Our project signifies a major step forward in making communication more accessible, especially for people with hearing difficulties. By incorporating advanced techniques from machine learning and machine vision, our system excels accuracy in converting sign language gestures into text and speech instantly. This precise interpretation improves communication access and fosters inclusivity for individuals facing hearing challenges. Through ongoing improvements and widespread use, our system has the significantly improve the quality of life for many individuals with hearing impairments, enabling them to communicate more confidently and independently in diverse social and professional environments. It's crucial to recognize certain constraints, including the necessity for additional fine-tuning and streamlining to improve accuracy and resilience, especially in intricate and evolving surroundings. Furthermore, relying solely on one language model could limit the system's adaptability to various linguistic contexts, underscoring the significance of integrating multilingual support in future iterations. This acknowledgement emphasizes the ongoing need for refinement and expansion to ensure broader applicability and effectiveness. Furthermore, considerations regarding user interface design and accessibility should be addressed to ensure seamless integration and usability for individuals with varying degrees of hearing impairment. We analyze the primary discoveries and ramifications of our endeavor, along with possible restrictions and pathways for future investigations. Our system's effective transformation of sign language hand gestures into written and spoken words highlights its capacity to notably enhance communication accessibility for people with hard of hearing challenges. Through the utilization of machine learning and computer vision methods, our system showcases encouraging precision and efficiency, establishing a solid foundation for applicable uses in real-life scenarios.

## VII. CONCLUSION

Our project signifies a major step forward in making communication more accessible, especially for people with hard of hearing difficulties. By incorporating advanced techniques from machine learning and computer vision, our system achieves high accuracy in translating sign language hand gestures into text and speech instantly. This precise interpretation improves communication access and fosters inclusivity for individuals facing hearing challenges. Through ongoing improvements and widespread use, our system has the capacity to greatly positively impact the well-being of individuals with hearing impairments, enabling them to communicate more confidently and independently in diverse social and professional environments.

## VIII. FUTURE SCOPE

In future development, expanding the system's capabilities by incorporating a larger and diverse data set would enhance performance and generalization. Additionally, integrating support for multiple languages would broaden accessibility, enabling accurate interpretation and conversion of sign language across different languages. This expansion would facilitate global communication for individuals experiencing hearing limitations these enhancements represent crucial steps in advancing ensuring inclusiveness in communication technology.

## REFERENCES

- [1] Mrs. Arepula Swetha, Vamja Pooja, Vundi Vedavyas, Challa Datha Vekata Naga Sai Kiran, Sadu Sravani. Sign Language to Speech Translation using machine Learning. Journal of Engineering Sciences (2022)
- [2] Punekar, A.C.Nambiar, S.A.Laliya, R.S, & Salian, S.R. (2020). A Translator for Indian Sign Language to Text and Speech. Int.J.Res.Appl.Sci.Eng.Technol.8,1640- 1646.
- [3] Yadav, B. K. Jadhav, D.Bohra, H, & Jain, R . (2021). Sign Language to Text and Speech Conversion. International Journal of Advance Research, Ideas and Innovations in Technology. [www.IJARIT.com](http://www.IJARIT.com)
- [4] Raskar, S. , Dahibhate, S., Yadav, A, & Randhe, K. Conversion of sign Language to text and speech. International Research Journal of Modernization in Engineering Technology and Science (IRJMETS) (2022)
- [5] Akshatharani, Bharath Kumar, and N.Manjanaik, Sign Language to text-speech translation using machine learning. International Journal of Emerging Trends in Engineering Research 9.7 (2021).
- [6] Prabhakar, Medhini, Prasad Hundekar, Sai Deepthi BP, and Shivam Tiwari. Sign Language Conversion to Text and Speech. (2022).
- [7] Viswanathan, S. Pandey, S. Sharma, K, & Vijayakumar, P. (2021). Sign language to text and speech conversion using CNN. International Research Journal of Modernization in Engineering Technology and Science (IRJMETS).
- [8] Adewale, V, & Olamiti, A. (2018). Conversion of sign language to text and speech using machine learning techniques. Journal of research and review in science. 5(12), 58-65.
- [9] Ojha, A., Pandey, A., Maurya, S., Thakur, A., & Dayanda, P. (2020). Sign language to text and speech translation in real time using convolution neural network. Int. J. Eng. Res. Technol. (IJERT), 8(15), 191-196.
- [10] Gore, S., Salvi, N., & Singh, S. (2022). Conversion of Sign Language into Text using Machine Learning Technique. International Journal of Research in Engineering, Science and Management, 4(5), 126-128



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)