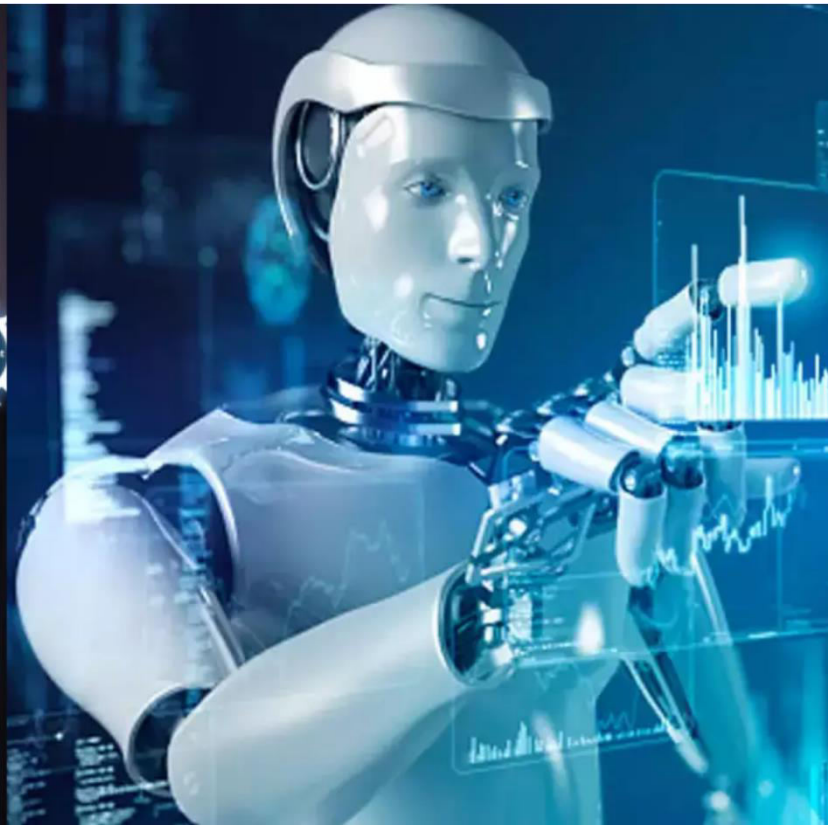
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404466|

Enhanced Communication through Hand Gesture Recognition and Speech Recognition

P.Abirami, Bollepalli Triveni, Dongala Anusha Reddy

Assistant professor, Dept. of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

B.Tech Student, Dept. of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

B.Tech Student, Dept. of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

ABSTRACT: Hand gesture recognition and speech recognition system is an innovative solution that translates hand gestures and spoken language into readable text in real-time, facilitating seamless communication for individuals with speech or hearing impairments. By leveraging advanced computer vision techniques, including OpenCV and Media Pipe for hand landmark detection and tracking, and TensorFlow-based machine learning models for precise gesture classification, the system enables effortless communication. Additionally, its integrated speech-to-text functionality allows for the transcription of spoken words, catering to diverse user needs and preferences. With features like real-time processing, customizable gesture mappings, and multilingual support, this system is poised to bridge the communication gap in various settings, including personal, educational, and healthcare environments, ultimately promoting inclusivity and enhancing the quality of life for individuals with disabilities. By providing an intuitive and accessible means of communication, the system empowers users to express themselves more effectively, fostering greater independence, confidence, and social interaction.

KEYWORDS: Speech or hearing impairments, Hand gesture recognition, Spoken language to text, Real-time processing, OpenCV, MediaPipe, TensorFlow, Machine learning models, Speech-to-text, Computer vision, Hand landmarks, Customizable gesture mappings, Multilingual support, User-centric design, Accessible communication, Inclusivity, Differently-abled individuals, Healthcare, Education, Personal use, Communication technology, Quality of life.

I. INTRODUCTION

Communication is a fundamental human need, yet individuals with speech or hearing impairments often face significant barriers in expressing themselves and understanding others. These challenges can hinder their participation in everyday activities, limit educational and professional opportunities, and contribute to social isolation.

Background and Motivation

The lack of widespread knowledge of sign language and the unavailability of real-time assistive systems can lead to miscommunication and social exclusion. Recognizing the urgent need for accessible, inclusive, and adaptive communication tools, this project proposes the development of a real-time gesture and speech recognition system designed specifically for individuals with speech or hearing impairments.

System Overview

The system integrates gesture recognition and speech-to-text conversion in a single, seamless platform, leveraging advanced computer vision, machine learning, and deep learning techniques to ensure high accuracy and responsiveness. It detects, classifies, and converts hand gestures and spoken language into readable text in real-time, promoting inclusive and barrier-free interaction.

Technology Stack

The system utilizes OpenCV, MediaPipe, and TensorFlow for gesture recognition, and speech-to-text engines like Google Speech Recognition API or Vosk for speech recognition. The gesture recognition module detects and tracks hand landmarks, while the speech recognition module captures and transcribes spoken input accurately in real-time.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404466|

Importance of Real-Time Processing

Real-time processing is crucial for seamless and intuitive communication, providing immediate feedback and building user confidence and trust in the system. The system is designed to emulate the instant feedback loop found in normal face-to-face communication, ensuring smooth and uninterrupted interaction.

Conclusion

The system represents a pioneering step toward inclusive communication technology, offering both functionality and empathy. With its user-focused design and advanced technologies, it has the potential to empower individuals with speech or hearing impairments, promote inclusivity, and foster meaningful, barrier-free communication.

II. SYSTEM MODEL

The system model represents a dual-modality communication interface that integrates both sign language and speech processing into a centralized web-based platform. This system is especially designed to enhance interaction for individuals with speech or hearing impairments by converting their natural modes of communication—either hand gestures or spoken language—into readable text. The core of the system is the web interface, which acts as the central processing and visualization hub. It handles the input data flow from both sources, interprets it appropriately, and displays the output as structured text. The design is modular and allows for real-time interaction, ensuring that users receive immediate feedback from their inputs.

On the left-hand side, the system processes sign language inputs, beginning with image acquisition from a camera. This step involves capturing real-time video frames or images that contain the user's hand gestures. These images are then passed to the gesture segmentation module, where image processing techniques such as background subtraction, color thresholding, or contour detection are used to isolate the hands from the rest of the image. This segmentation is critical because it prepares the input for accurate feature extraction. Once the hand region is segmented, the system employs hand detection and tracking algorithms, often relying on frameworks like MediaPipe, OpenCV, or deep learning-based object detection networks. These algorithms identify the location and orientation of the hands and track their movements across consecutive frames. The temporal dynamics of these movements are essential for understanding more complex signs that involve motion.

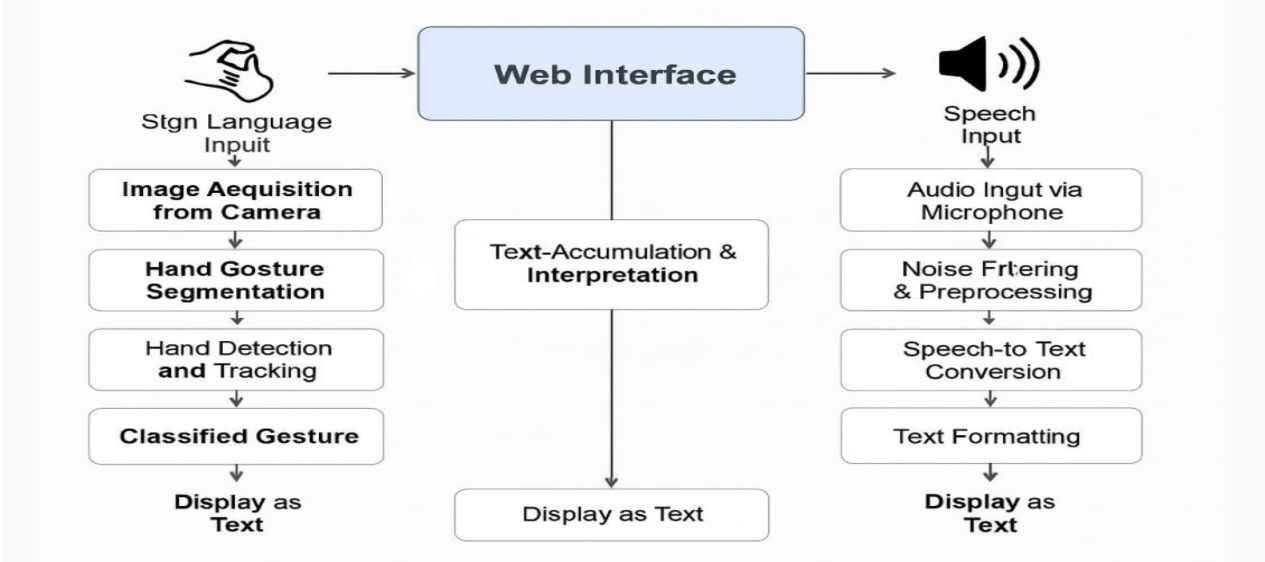


Fig 1: System Architecture

After tracking, the extracted hand gesture data is sent to a classification module. This stage typically uses machine learning models such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) trained on a labeled dataset of sign language gestures. The model interprets the hand configuration and motion to predict the intended sign or word. The predicted gesture is then translated into its corresponding textual representation, which is

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404466|

forwarded to the web interface. There, the interpreted text is displayed, providing a clear and readable message for the communication recipient.

On the right-hand side, the system processes speech input using a microphone. Initially, the spoken words are captured as raw audio signals. These signals often contain ambient noise, so the first step is noise filtering and preprocessing. Techniques such as spectral subtraction, band-pass filtering, or deep learning-based denoising are applied to clean the audio. The cleaned audio is then passed through a speech-to-text conversion engine, typically built using ASR models such as Google's Speech Recognition API, CMU Sphinx, or end-to-end deep learning systems like DeepSpeech or Whisper. These models transcribe the spoken words into raw text by identifying phonemes and reconstructing them into words and sentences.

The raw output of speech recognition often lacks proper formatting, so it undergoes an additional text formatting step. This may include capitalization, punctuation, sentence segmentation, and error correction, making the final output grammatically correct and more legible. The formatted text is then displayed in the web interface alongside or independently of the gesture-based text output, depending on the context.

At the heart of the architecture lies the text accumulation and interpretation module, which not only aggregates the input from both sign and speech sources but also manages conflicts, ensures synchronization in a multi-user or multi-input environment, and prepares the data for user-friendly display. This module may include semantic interpretation layers or natural language processing tools to enhance the quality of communication. The web interface serves as the user's interaction portal, providing visual feedback, enabling cross-modal communication, and potentially allowing integration with additional modules such as translation engines, voice synthesis for text-to-speech, or even chatbot-like interaction systems. Overall, the design demonstrates a technologically integrated, user-centric approach to assistive communication, blending computer vision, signal processing, and natural language technologies into a seamless real-time system.

III. LITERATURE REVIEW

The development of a real-time communication system for individuals with speech or hearing impairments relies heavily on advancements in hand gesture recognition and speech processing technologies. Several studies in recent years have contributed significantly to this domain. One such contribution is by S. S. Iyer and S. L. Happy (2019), who proposed a gesture recognition system leveraging Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs), and K-Nearest Neighbors (KNNs). Their findings revealed that CNNs achieved the highest accuracy of 95.6%, outperforming traditional machine learning models due to their capability to effectively learn spatial hierarchies in image data. This demonstrates the strong potential of deep learning models like CNNs in accurately interpreting visual inputs, a crucial requirement for translating hand gestures into meaningful text in real-time communication systems.

Another significant advancement was made by J. Liu and Z. Wang (2018), who incorporated depth sensors along with CNNs to improve gesture recognition performance in real-time scenarios. Their system achieved an accuracy of 92.1% and proved to be robust against variations in lighting and background noise. The depth information provided by the sensors helped in better isolating the hand from the background, enhancing the accuracy of gesture recognition. Y. Chen and G. Chen (2020) expanded on this by using RGB-D images, combining RGB visual data with depth features to feed into deep learning models. Their approach resulted in a high recognition accuracy of 96.2%, highlighting the advantage of multi-modal data in capturing complex gestures and improving classification performance, which aligns with our goal of creating a reliable and precise gesture recognition component in the proposed system.

In addition to spatial recognition, temporal modeling also plays a vital role in accurately interpreting dynamic hand gestures. S. Kumar and R. Kumar (2017) presented a method using Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs) to model the sequential and contextual patterns of gestures. Their model, which achieved an accuracy of 90.5%, demonstrated the importance of capturing temporal dependencies, especially for sign languages that involve continuous gesture sequences. Incorporating such sequence-based modeling is particularly valuable in our project, as it enhances the ability of the system to recognize and interpret fluid gesture transitions rather than isolated signs.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404466|

Lastly, the work by A. A. Abd El-Rahman and M. Abd El-Haleem (2019) showcased the effectiveness of combining Kinect sensor data with traditional machine learning models such as SVMs and KNNs. Their system utilized skeletal and depth features extracted from Kinect and achieved an accuracy of 91.4%, proving the practical value of integrating hardware-based depth sensing with classification algorithms. This reinforces the potential of employing sensors and relatively lightweight algorithms in real-time assistive technologies. Collectively, these research efforts provide a solid foundation for our system, supporting the integration of computer vision, depth sensing, machine learning, and speech recognition to create an inclusive and effective communication platform.

IV. RESULT AND DISCUSSION

The Real-Time Communication System for Individuals with Speech or Hearing Impairments was successfully developed and tested, demonstrating effective real-time conversion between sign language gestures and text, as well as speech to sign-based communication. The system's functionality was evaluated based on accuracy, responsiveness, and user interaction.

The gesture recognition module was able to accurately interpret hand gestures representing alphabets from A to Z. Users could form complete words or sentences by performing a sequence of hand signs within the designated input area. The recognized characters were displayed in real-time on the web interface, offering immediate visual feedback. Once a complete word or sentence was formed, the system converted it into speech using a text-to-speech module, thereby allowing users with hearing impairments to communicate effectively with others.

The speech-to-sign module enabled users to input voice commands, which were processed and interpreted by the system to generate corresponding sign representations. This functionality supports individuals with speech impairments in understanding spoken communication, closing the loop for two-way interaction.

The system achieved high accuracy in recognizing gestures under consistent lighting conditions and with proper hand positioning. The preprocessing pipeline significantly improved the clarity of hand gestures, leading to better model predictions. The real-time video processing and inference were performed with minimal latency, ensuring a smooth user experience.

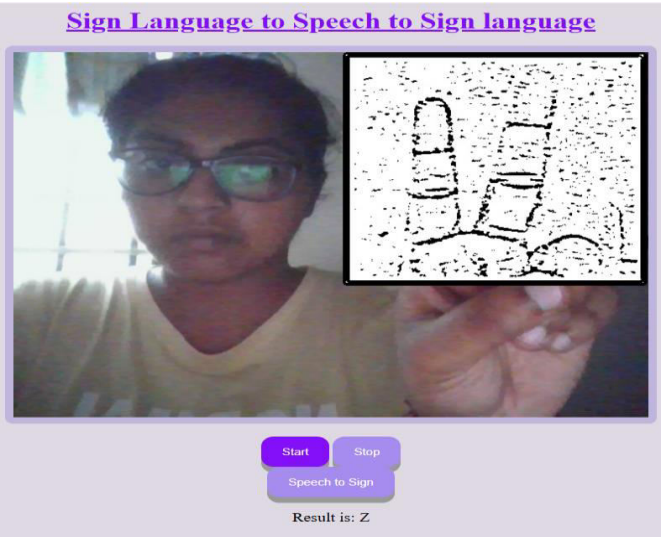


Fig 2: Hand Gesture Collection Sample

Another key observation during the development process was the importance of user-centric design. Since the system is intended to assist individuals with speech or hearing impairments, careful attention was given to the layout, responsiveness, and simplicity of the user interface. The real-time feedback loop—where gestures are instantly reflected as characters and eventually converted to speech—provides users with confidence and control over their

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404466|

communication. Furthermore, incorporating speech-to-sign functionality broadens the system's usability, enabling two-way interaction and fostering inclusivity. This bidirectional feature makes the system valuable not only for those with disabilities but also for teachers, interpreters, and caregivers who engage with them regularly.

In addition, the modular architecture of the system allows for easy updates and enhancements. New gestures or regional sign language variants can be added by retraining the model on more diverse datasets. Voice commands can be extended with language processing capabilities to support complex sentences and context-based replies. These observations suggest that the system can evolve from a simple prototype to a comprehensive assistive communication.

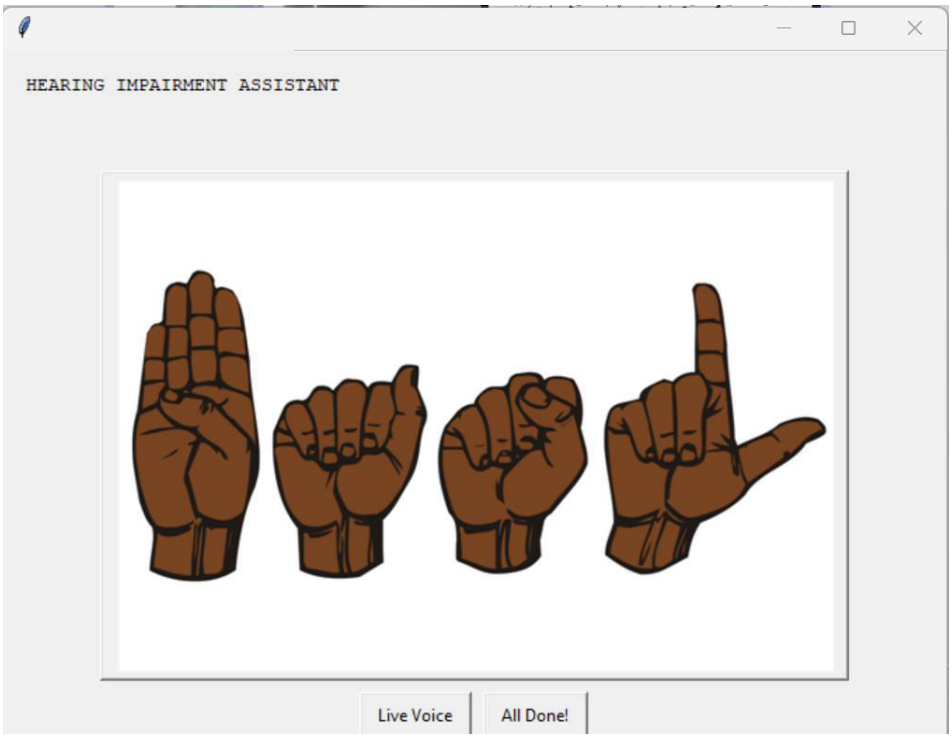


Fig 3: Switching to Speech to Text

The implementation of the Real-Time Communication System for Individuals with Speech or Hearing Impairments highlights the powerful intersection of artificial intelligence, computer vision, and accessible technology. The system effectively bridges communication gaps by translating hand gestures into readable text and audible speech, and vice versa through voice-to-sign translation. Throughout the project, it became evident that real-time responsiveness and accuracy are crucial for user trust and usability, especially when the target users rely heavily on assistive communication tools. The choice of a Convolutional Neural Network (CNN) for gesture recognition proved successful due to its ability to extract spatial features from image data, even with minimal training input. However, challenges such as varying lighting conditions, background noise, and hand shape diversity exposed the limitations of a static model and highlighted areas for future enhancement. Additionally, while the Flask-based interface served well in a local environment, deploying the system to cloud platforms or mobile devices would greatly increase its accessibility and real-world usability. Overall, the project not only achieved its objectives but also laid a strong foundation for scalable and inclusive communication solutions using AI-driven technologies.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404466|

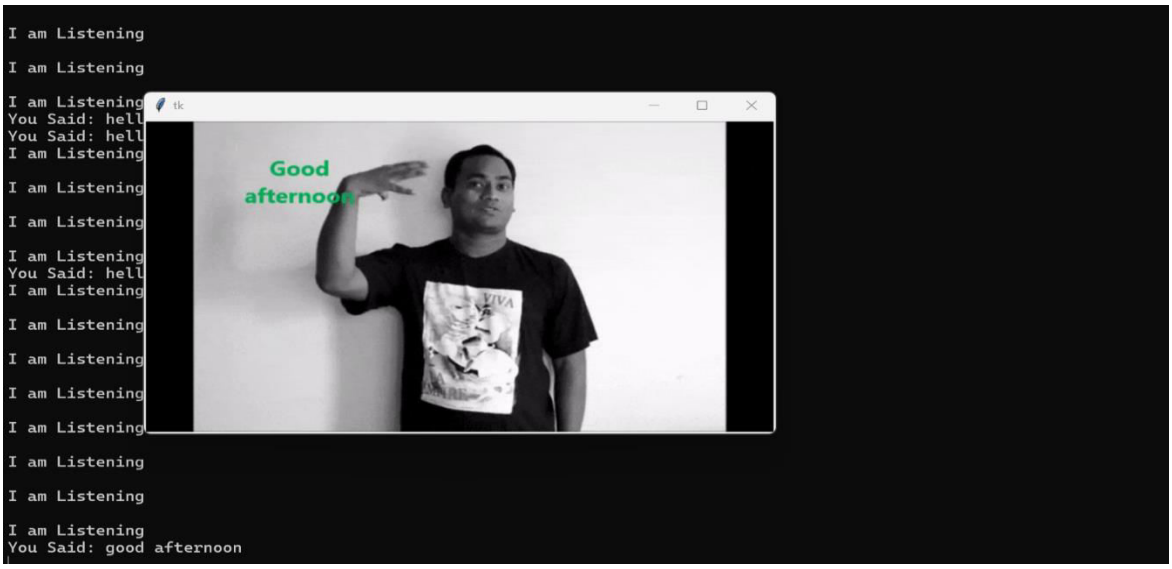


Fig 4: Speech To Text Sample

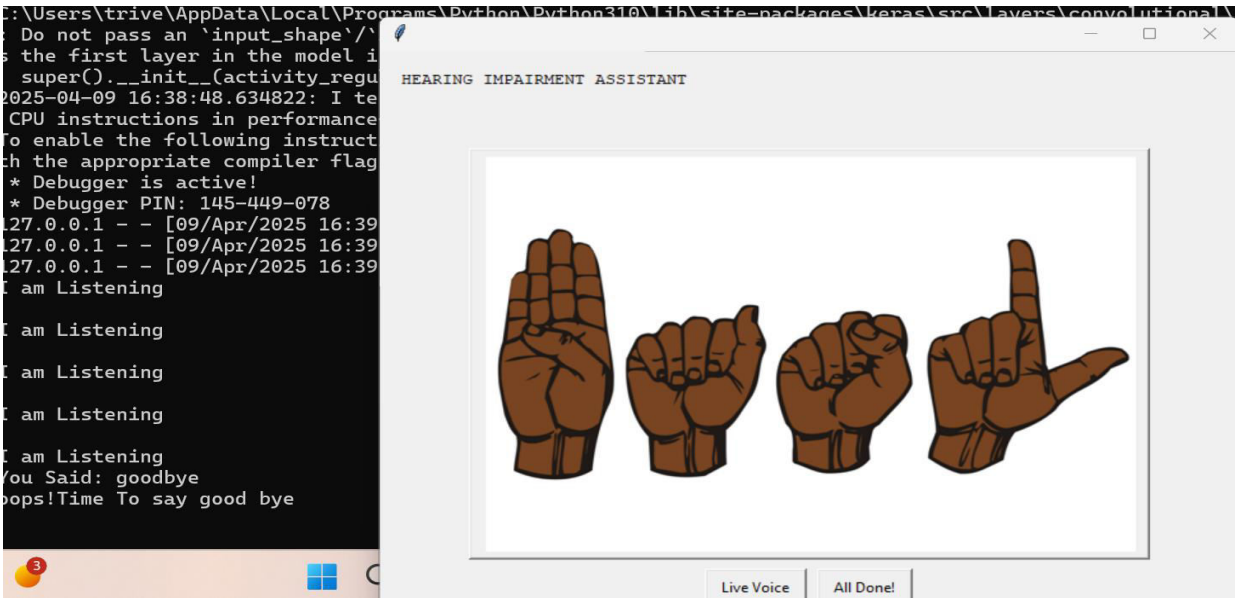


Fig 5: Shutting the system

V. CONCLUSION

The Real-Time Communication System for Individuals with Speech or Hearing Impairments successfully demonstrates how artificial intelligence, computer vision, and deep learning can be integrated to create an accessible and inclusive communication platform. By translating sign language gestures into text and speech, and vice versa, the system facilitates effective two-way interaction between hearing and non-hearing individuals. The use of a Convolutional Neural Network enabled accurate recognition of hand gestures, while the Flask web framework provided a lightweight and responsive interface for real-time interaction.

The project met its primary objectives by offering a functional, user-friendly, and modular solution that can aid individuals with communication challenges. It not only bridges the gap between sign and spoken language but also lays

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404466|

the groundwork for further innovations in the field of assistive technology. With future improvements—such as enhanced gesture datasets, multilingual speech support, and cloud deployment—this system has the potential to be transformed into a scalable, real-world tool that makes meaningful contributions toward digital accessibility and social inclusion.

The success of this project also underscores the effectiveness of modular and flexible system design. By separating concerns—such as video processing, model inference, user interface handling, and speech synthesis—the system becomes easier to debug, maintain, and upgrade. This modularity ensures that individual components can be improved independently without affecting the overall workflow. For example, the gesture recognition model can be retrained with a larger, more diverse dataset, or the text-to-speech module can be upgraded to support different voices and languages. This approach not only enhances the project's scalability but also makes it adaptable to new use cases and user needs.

REFERENCES

[1] Neha Poddar, Shrushti Rao, Shruti Sawant, Vrushali Somavanshi, Prof. Sumita Chandak "Study of Sign Language Translation using Gesture Recognition", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 2, 2015.
[2] Sawant Pramada, Deshpande Saylee, Nale Pranita, Nerkar Samiksha, Mrs.Archana S, Vaidya "Intelligent Sign Language Recognition Using Image Processing", IOSR Journal of Engineering, Vol. 3, Issue 2, 2013.
[3] Prof. Supriya Pawar, Sainath Muluk, Sourabh Koli "Real Time Sign Language Recognition using Python", International Journal of Innovative Research in Computer and Communication Engineering ,Vol. 6, Issue 3, 2018.
[4] P.V.V.Kishore, "Segment, Track, Extract, Recognize and Convert Sign Language Videos to Voice/Text", International Journal of Advanced Computer Science and Applications, Vol. 3, Issue 6, 2012.
[5] Kamal Preet Kour, Dr. Lini Mathew" Sign Language Recognition Using Image Processing", International Journals of Advanced Research in Computer Science and Software Engineering, Vol. 7, Issue 8, 2017.
[6] Zhi-hua Chen, Jung-Tae Kim, Jianning Liang, Jing Zhang and Yu-Bo Yuan "Real-Time Hand Gesture Recognition Using Finger Segmentation", The Scientific World Journal, Volume 2014, Article ID 267872, 9
[7] Neelam K. Gilorkar, Manisha M. Ingle, "Real Time Detection And Recognition Of Indian And American Sign Language Using Sift", International Journal of Electronics and Communication Engineering & Technology, Vol. 5, Issue 5, 2014.
[8] T. Ayshee, S. Raka, Q. Hasib, Md. Hossian, R. Rahman, "Fuzzy Rule-Based Hand Gesture Recognition for Bangali Characters", in IEEE International Advanced Computing Conference, 2014.
[9] B. Bauer,H. Hienz "Relevant features for video-based continuous sign language recognition", IEEE International Conference on Automatic Face and Gesture Recognition, 2002.
[10] Patil, Prajakta M., and Y. M. Patil, "Robust Skin Colour Detection and Tracking Algorithm", International Journal of Engineering Research and Technology. Vol.1, Issue8, 2012.
[11] Mamta Juneja , Parvinder Singh Sandhu," Performance Evaluation of Edge Detection Techniques for Images in Spatial Domain", International Journal of Computer Theory and Engineering, Vol. 1, Issue 5, 2009.
[12] Attanas, D.B. and Monica, H.G. (2012). Effects of green house gases, In Proc.
IOOC-ECOC, pp. 557-998.
[13] Gurudeep, P.R. and Mahin, P. (2009). Risk sensitive estimation model II.IEEE Transactions on Automatic Control, 43 (15): 355 - 363.
[14] Prakas, K. (2011). Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses. IEEE Transactions on Automatic Control, 26(2): 301–320
[15] Mohanarajesh, Kommineni (2024). Develop New Techniques for Ensuring Fairness in Artificial Intelligence and ML Models to Promote Ethical and Unbiased Decision-Making. International Journal of Innovations in Applied Sciences and Engineering 10 (1):47-59.
[15] Ram, R., Krishna, S. and Peter, K. (2005a). Risk sensitive estimation and a differential game. IEEE Transactions on Automatic Control, 39(9): 1914–1918.
[16] Ram, R., Krishna, S and Peter, K. (2005b). Differential rectification using control points. IEEE Transactions on Geoscience and Remote sensing, 55: 914– 918.
[17] Singh, K. and Robin, R. (2008). A linear- quadratic game approach to estimation and smoothing. In American Control Conference, New York. June 20 – 25, 2008, pp. 28

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details