International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 12 | Dec - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Connecting Sign Language and Speech with Multilingual Translation Using Random Forest

[1]Selva Sheela. K, [2]Praveena. R, [3]Sivaramana. H, [4]Rujuta.M, [5]Vaishnavi.R
[1] Assistant Professor, [2,3,4,5]Student
[1,2,3,4,5] Department of Artificial Intelligence and Data Science,
[1,2,3,4,5] KGiSL Institute of Technology, Coimbatore, Tamil Nadu

Abstract
This project demonstrates an innovative web application that aims to deepen the communication between gesture based clients and users of communicating through language. The application relies on advanced PC vision and AI algorithms for a seamless interpretation between gesture and voice communications. By including MediaPipe for hand gesture recognition and the Irregular Backwoods computation for movement classification, the system can understand and translate motions into equivalent words in Tamil and English. The app makes use of a webcam that captures hand movements, feeds them through a pre-trained Irregular Backwoods model, and then produces the interpreted gesture-based communication in text form. The constant nature of the application allows clients to share easily and multilingual support further fosters openness. The model's ability to be physically prepared and changed ensures versatility and accuracy in a vast range of settings. With this innovation, the drive aims to fill communication gaps and increase accessibility for the deaf and hard of hearing community, along with overcoming the limitations of the current sign language interpretation systems.
Keywords: Real-time Sign Language Translation 1, Machine Learning 2, Random Forest 3, MediaPipe 4, Multilingual Support 5.

1. INTRODUCTION

In an era where digital solutions are progressively addressing communication hurdles, the demand for accessible and inclusive technology has never been greater. This project provides a cutting-edge online tool that enables real-time communication between sign language users and people who use spoken language. This program addresses the special issues faced by the deaf and hard-of-hearing community by offering an efficient and user-friendly platform for translating sign language into voice and vice-versa.

The application makes use of modern technologies such as MediaPipe for precise hand gesture recognition and the Random Forest algorithm for accurate gesture classification. By combining these technologies, the system can understand and convert hand gestures into written and spoken words, including Tamil and English translations.

The application features a simple three-phase workflow. Initially, users capture their hand motions with a webcam, which are then processed and evaluated by the software. The pre-trained Random Forest model categorizes the motions, and the system converts them into the related words or sentences. This real-time translating capability allows users to engage in meaningful conversations without delays or misinterpretations.

Notwithstanding motion acknowledgment and interpretation, the application accommodates manual model preparation to deal with different developments and sign varieties. This flexibility guarantees that the framework is precise and valuable across many gesture based communications and vernaculars. The joining of this innovation tends to different disadvantages of customary communication through signing deciphering frameworks, including the prerequisite for human translators and the unconventionality in communication through signing aptitude. Via robotizing the interpretation interaction and offering language support, the program further develops correspondence productivity and inclusivity.

The significant reason for this undertaking is to foster a device that increments correspondence access for the hard of hearing and deaf local area, consequently making collaborations with the bigger society more comprehensive and productive. The application's utilization of cutting edge computer based intelligence calculations and continuous handling guarantees that it serves the necessities of its clients while likewise pushing the limits of communication through signing interpretation innovation.

2. LITERATURE SURVEY

Using technologies like OpenCV, MediaPipe, Convolutional Neural Networks (CNNs), scikit-learn, and various APIs for text-to-speech conversion, a thorough literature review for a project involving real-time sign language to speech and speech to sign language translation reveals a rich landscape of research and technological advancements in computer vision, machine learning, and real-time processing.

Recent literature has thoroughly examined the integration of computer vision and machine learning in gesture recognition systems. Google's MediaPipe is a well-known technology for gesture recognition and real-time hand tracking. It is a useful tool for translating sign language movements into text and speech because it provides a strong framework for accurately identifying and interpreting hand landmarks.Numerous studies that emphasize MediaPipe's capacity to provide precise and effective hand tracking have shown how successful it is in real-time applications [1][2].

Convolutional Neural Networks (CNNs) are gaining a lot of attention in the machine learning community due to their impressive performance in image and gesture recognition tasks. CNNs are well known for their capacity to extract hierarchical information from pictures, which is crucial for deciphering intricate sign language gestures. Although the Random Forest algorithm is the primary method used in this study for classifying gestures, CNNs have shown promise in related fields and could lead to increased performance and accuracy in gesture detection systems [3][4].

The literature has extensively documented the use of Random Forest, a flexible ensemble learning technique, for hand gesture classification. This technique is appropriate for jobs involving gesture classification since it can handle high-dimensional data and combine predictions from several decision trees. Random Forest is a dependable option for this project since research indicates that it can classify different hand motions with great accuracy [5][6].

For Computer Vision undertakings OpenCV is a famous library that is fundamental for ongoing picture handling and representation it offers works that are crucial for executing the continuous piece of communication through a signing interpretation framework, for example, recording video outlines handling pictures, and introducing results the library's wide application in PC vision applications features how well it handles video transfers and controls pictures.

Devices like Pickle which is utilized to save and load AI models assist with information serialization and model organization pickles ability to serialize Python objects ensures the viable putting away and recovery of models thus supporting the execution and upkeep of AI frameworks.
A few APIs like Microsoft purplish blue mental administrations and Google Cloud text-to-voice give dependable text-to-voice transformation benefits that produce a discourse that seems normal these APIs supplement the venture multilingual help by supporting different dialects and voices the writing additionally stresses how significant multilingual help is to further developing client openness and contribution innovation arrangements that consolidate language support ensure that an assortment of client gatherings can draw in with and benefit from the application in a proficient way.

Everything considered the usage of these technologies text-to-discourse APIs for voice yield CNNs and irregular timberland for motion recognition media pipe for hand following OpenCV for ongoing handling and OpenCV for constant processing shows a careful way to deal with making a fruitful communication via gestures interpretation framework by tending to correspondence leaps the blend of these instruments offers clients a compelling and comprehensive arrangement.

3. PROPOSED SYSTEM

The suggested system intends to overcome the constraints of existing models by proposing a multilingual sign language translation system that can convert sign language into text and speech in both English and Tamil. The current technology aims to facilitate real-time communication between the hearing and speech-impaired communities and the rest of society. This solution is designed to improve communication efficiency and inclusivity by offering a comprehensive, user-friendly interface that can be expanded to support additional languages in the future.

The system recognizes hand movements using a combination of powerful machine learning methods, including Convolutional Neural Networks (CNNs) and the Random Forest classifier. Real-time hand tracking and image processing are enabled by the integration of OpenCV and MediaPipe, resulting in accurate recognition and classification of sign language hand gestures. The model is trained to detect individual letters and popular words in both English and Tamil, allowing users to communicate in their favorite language.

Furthermore, the use of scikit-learn and pickle allows for efficient maintenance and storage of trained models, guaranteeing that the system can be continuously retrained and updated as new gestures or phrases are introduced. The system is designed to be scalable, allowing for future development into additional languages and dialects.

For output, the project includes text-to-speech APIs, which allow the system to translate identified gestures into spoken words in real time. This improves user interaction, particularly among non-sign language users. The system's seamless transition between English and Tamil allows users to chose their preferred language, making it versatile and inclusive of all populations. The user interface is intuitive and responsive, making it easy to use by anyone with diverse levels of technical expertise. The interface ensures a smooth experience for both the sign language user and the person receiving the translated output, resulting in prompt, accurate communication.

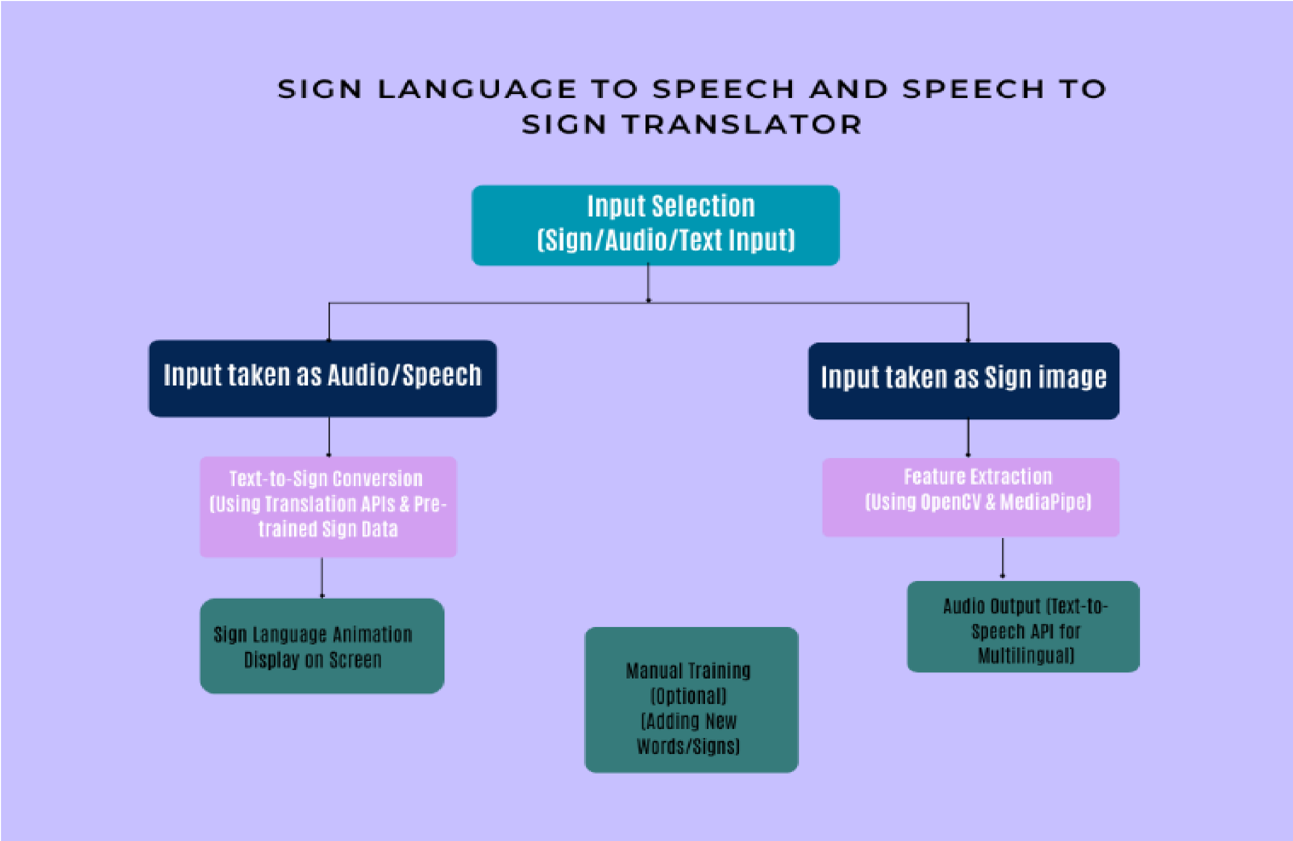

Fig. 1. Workflow diagram

This proposed technology reduces the barriers that sign language users frequently face by providing a real-time, multilingual solution. It is intended to increase social inclusion and facilitate more fluid interaction between sign language users and the larger community.

4. SYSTEM IMPLEMENTATION

During the system implementation phase, the project's conceptual design is transformed into a fully functional system. This step is crucial for ensuring that the system meets user needs and runs well in real-world circumstances. The proposed sign language translation system's major implementation components revolve around the seamless conversion of sign language motions to both text and speech, as well as vice versa, with an emphasis on user-friendliness.

The system's first key component is Sign Language to Text/Audio Conversion. Users can go to the website and utilize their webcam to sign words. When the user finishes their sentence and makes the "Thank You" sign, the system interprets the signs and outputs the entire sentence as audio and text. Hand motions are recognized using OpenCV and MediaPipe for hand tracking, while CNNs are employed for gesture classification into related words. The text is then turned into speech via a text-to-speech API, allowing the system to deliver audio feedback in real-time.
The Audio/Text to Sign Language Conversion function allows users to enter text or audio. The system interprets these inputs and converts them into corresponding sign language gestures that are presented on screen. This dual-mode input provides versatility by supporting both text and spoken words, making it suitable for a wide range of users.

Both capabilities do not require real-time training, increasing the system's speed and usability. However, a Manual Training Mode is available for users and developers to add additional phrases or signs to the system. This feature

enables the system to evolve dynamically, allowing new signals to be manually learned and added while updating both the sign language and the accompanying phrases in English or Tamilsimultaneously.
The implementation process was rigorously assessed using test data, and any problems were remedied in response to user feedback. The system also incorporates validation tests to verify data entry accuracy, as well as built-in error alerts to prevent improper sign recognition or translation. Training and educational support have been offered to ensure that users, regardless of technical knowledge, can navigate and utilize the system correctly.

Finally, the system has been thoroughly tested using both text and sign language inputs, yielding accurate results in multiple languages (English and Tamil). The manual training mode ensures that the system is constantly improved, making it an effective tool for bridging communication barriers between sign language users and the general public.

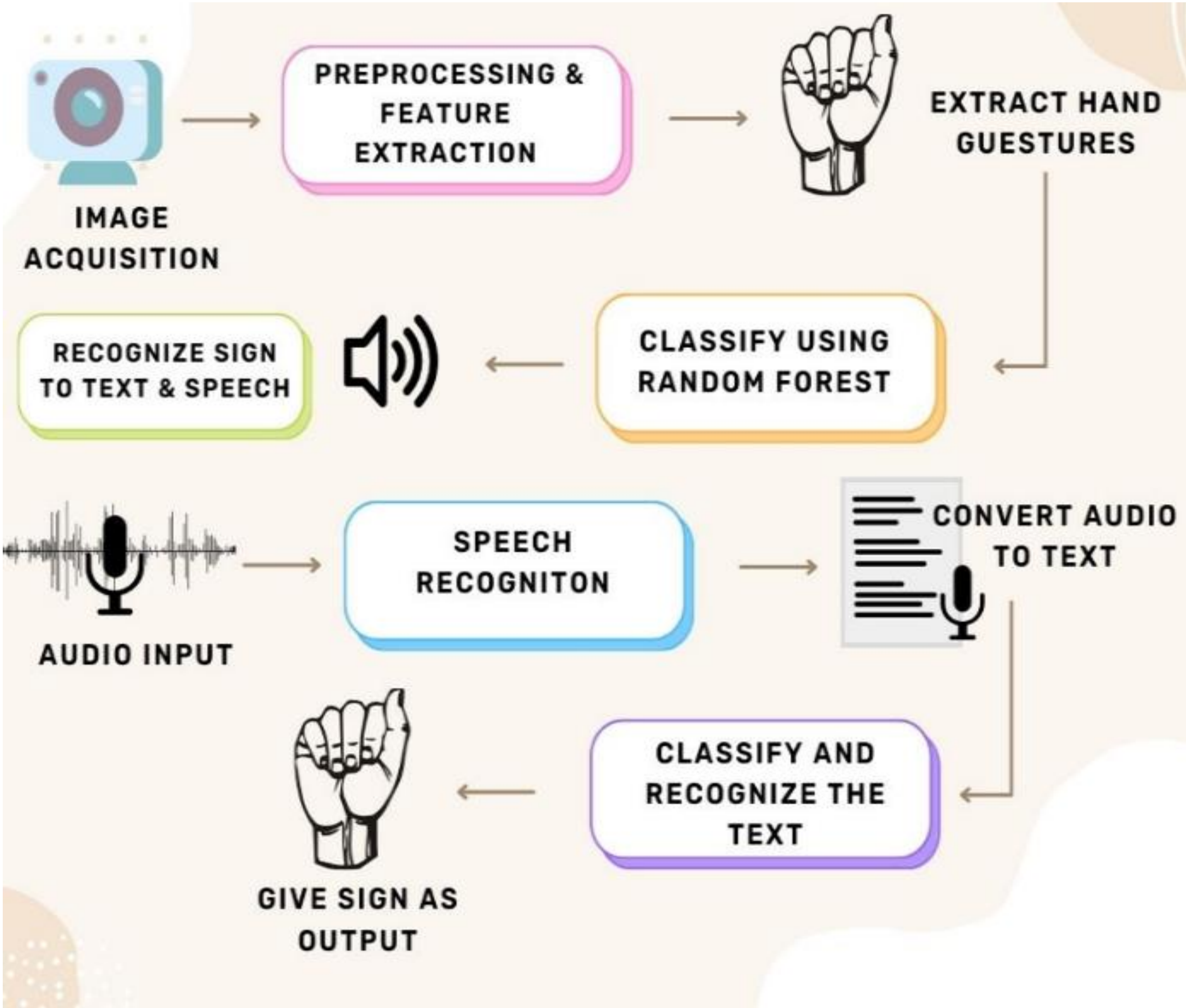

Fig. 2. Workflow diagram

5. RESULT

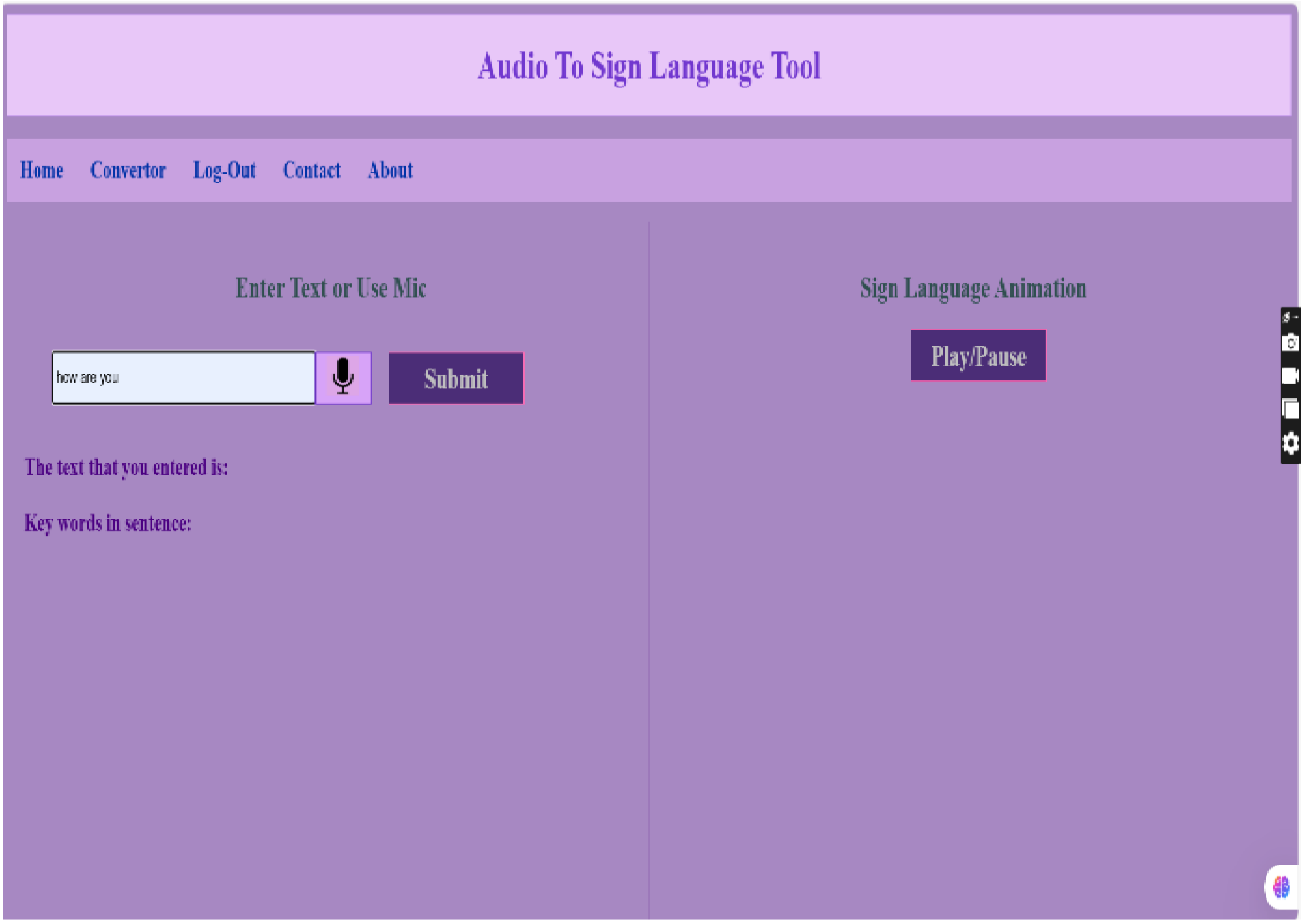

Fig. 3.

This image showcases a user-friendly interface where users can input a sentence using sign language. This input sentence then beprocessed for translation.

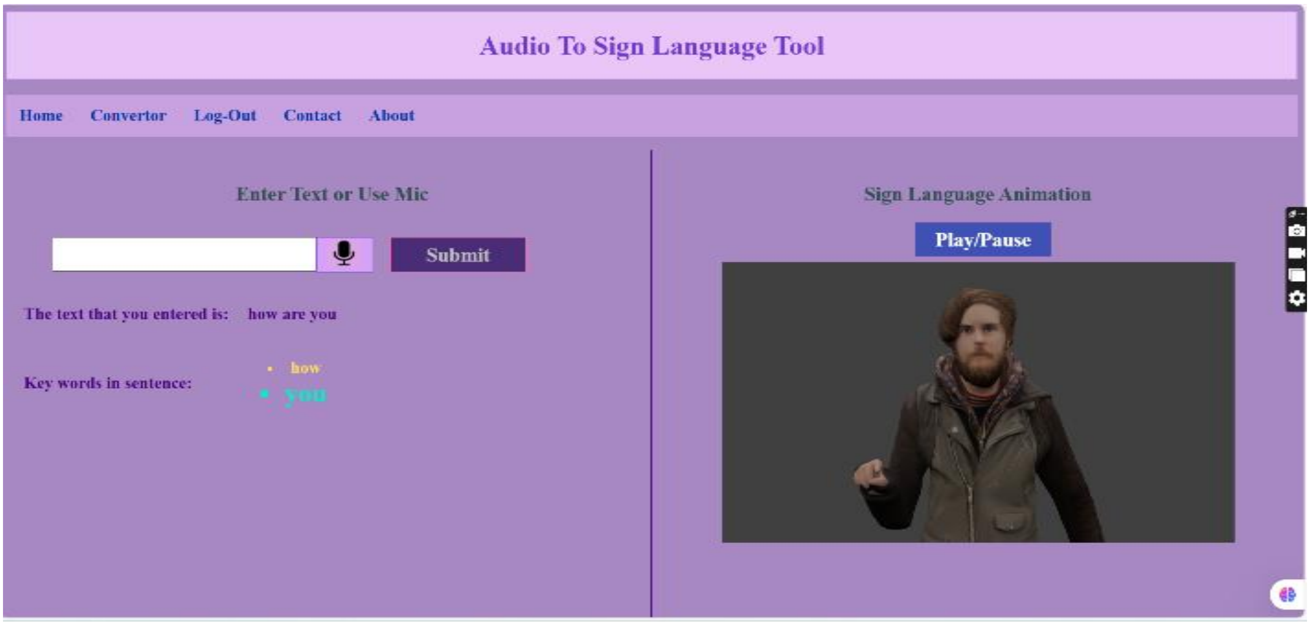

Fig. 4

This shows the generated signs corresponding to the input sentence, enabling users to learn and practice sign language. The video display enhances the learning experience, making it more engaging and interactive.

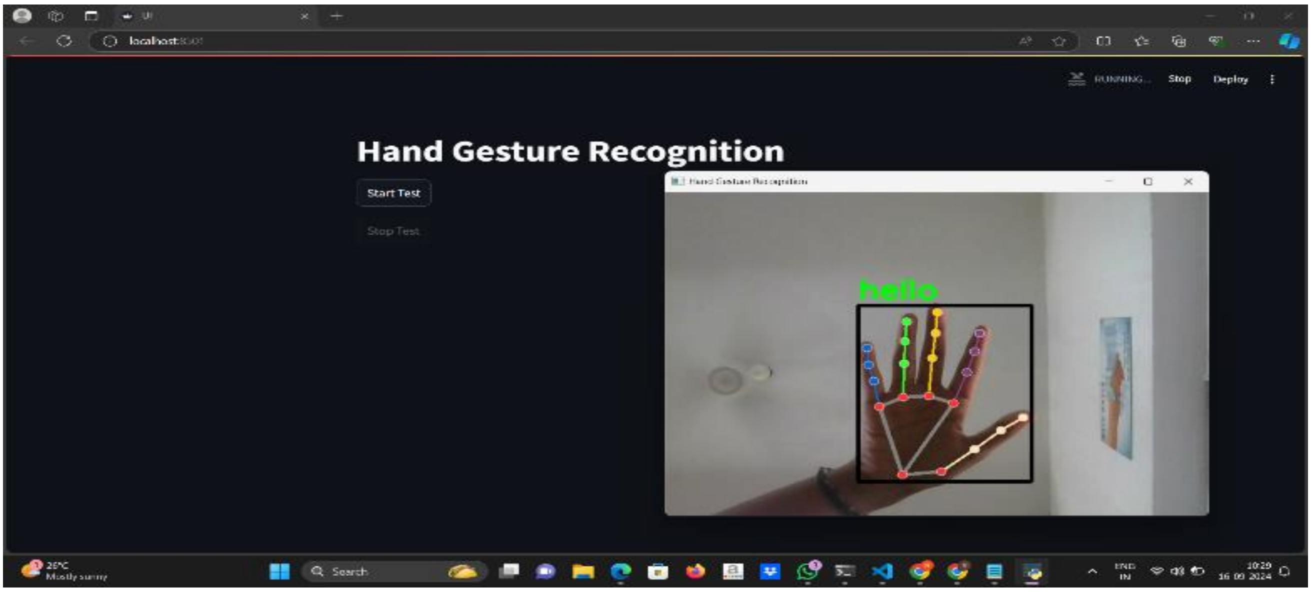

Fig. 5.

Our system uses computer vision to capture and analyze the user's hand gestures, detecting the input sign with high accuracy. The hand pinpoints are recognized and matched with a database of known signs. This enables the system to accurately identify and translate the input sign.

Fig. 6

The translated text of the input sign is displayed in the terminal, providing a clear and readable output. Simultaneously, the system plays an audio output of the translated text, enhancing accessibility for users with visual impairments. This feature enables users to verify the accuracy of the translation.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 12 | Dec - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

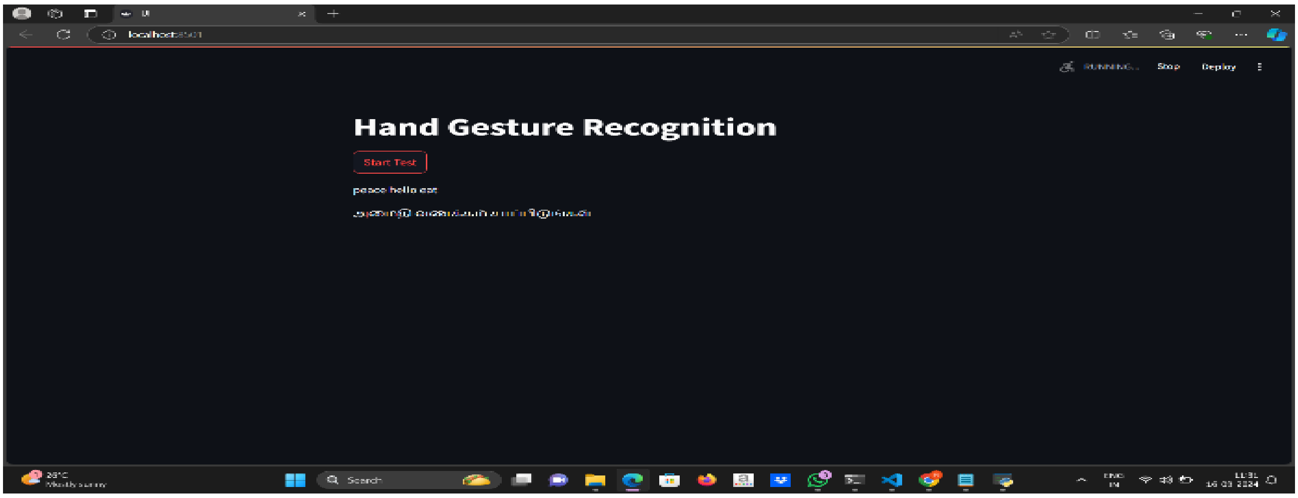

Fig. 7.

Our system seamlessly integrates language translation, enabling users to receive output in their preferred language. In this example, the translated text is displayed in the UI, and the audio output is played in Tamil, demonstrating the system's multilingual capabilities. This feature promotes inclusivity and accessibility for users with diverse language backgrounds.

6. CONCLUSION

This exploration offers a basic response for shutting the correspondence hole between the conference and discourse hindered networks and the overall population the framework gives an ongoing and multilingual interpretation administration for English and Tamil by transforming communication via gestures into text and sound as well as the other way around the expansion of sound to-sign interpretation and manual preparation capabilities expands the frameworks flexibility and versatility making it supportive to a different range of clients the major objective of the venture which was to advance consistent and exact correspondence was accomplished by incorporating trend setting innovations like OpenCV, CNN and text-to-discourse APIs the frameworks capacity to extend through manual preparation guarantees that it very well might be refreshed and changed in light of client needs as an electronic stage the innovation is basic and open permitting individuals to communicate really without the requirement for expert gear future advancements might incorporate expanded language support and further developed signal recognition the idea could be transformed into a cell phone application to build openness and make it more agreeable for everyday use the undertakings general points have been understood and the framework addresses a major step in the right direction in making correspondence simpler and more productive for those with hearing and discourse difficulties.

REFERENCE

[1]. Zisserman, A., & Szeliski, R. (2020). Computer Vision: Algorithms and Applications. Springer.

[2]. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[3]. Aggarwal, G., & Chaudhary, S. (2019). Sign Language Recognition System Using CNN and MediaPipe. International Journal of Computer Science and Mobile Computing.

[4]. Zhang, Z. (2012). Gesture recognition using depth-based hand tracking. IEEE Transactions on Multimedia.

[5]. Shukla, S., & Sharma, A. (2021). Machine Learning in Sign Language Recognition: A Survey. International Journal of Machine Learning and Computing.

[6]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[7]. Koller, O., Zargaran, S., Ney, H., & Bowden, R. (2016). Deep sign: Hybrid CNN-HMM for continuous sign

language recognition. Proceedings of the British Machine Vision Conference (BMVC).
[8]. L. Yu, M. Sun, and S. Wan. (2020). A Deep Learning Method for Sign Language Recognition Based on 3D Handshape Features. IEEE Access.
[9]. Rastgoo, R., Kiani, K., & Escalera, S. (2020). Sign language recognition: A deep survey. Expert Systems with Applications.
[10]. Baccouche, M., Mamalet, F., Wolf, C., Garcia, C., & Baskurt, A. (2011). Sequential deep learning for human action recognition. International Workshop on Human Behavior Understanding.
[11]. Moon, G., Chang, J. Y., & Lee, K. M. (2019). Camera distance-aware top-down approach for 3D multi-person pose estimation from a single RGB image. Proceedings of the IEEE International Conference on Computer Vision (ICCV).
[12]. Hossain, M. M., & Lee, S. H. (2019). Explaining hand gestures: Depth sensor-based gesture recognition for sign language translation. IEEE Transactions on Multimedia.
[13]. Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. Advances in Neural Information Processing Systems (NeurIPS).
[14]. Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[15]. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
[16]. Zhang, Y., & Tian, Y. (2015). RGB-D image-based sign language recognition using convolutional neural networks. Proceedings of the IEEE International Conference on Image Processing (ICIP).
[17]. Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[18]. Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters.
[19]. MediaPipe: Cross-platform, customizable ML solutions for live and streaming media. (n.d.). Retrieved from https://google.github.io/mediapipe/
[20]. OpenCV: Open Source Computer Vision Library. (n.d.). Retrieved from https://opencv.org/