ISSN: 2395-7852

International Journal of Advanced Research in Arts, Science, Engineering & Management

Volume 12, Issue 2, March- April 2025

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.028

Sign Lang Connect: Sign Language Recognition and Speech to Text Converter for Regional Languages

Mrs. R. Karthika[1], Sri Jothi Rekha. R[2]

Assistant Professor, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, India[1]

U.G Student, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, India[2]

ABSTRACT: The absence of widely available sign language translation tools, particularly for regional languages, continues to create communication barriers between the general public and the hearing and speech-impaired community. By launching a sophisticated two-way sign language converter designed for regional languages, Sign Lang Connect seeks to close this gap. In addition to translating spoken or written language into sign animations, this system uses deep learning and computer vision techniques to recognize and translate sign language gestures into text and speech. By integrating real-time gesture recognition with a customizable regional language database, Sign Lang *Connect* enhances accessibility, ensuring seamless and inclusive communication. The proposed model emphasizes accuracy, adaptability, and ease of use, making it a transformative tool for education, social interaction, and professional settings.

KEYWORDS: Sign Language Recognition, Speech-to-Text, Deep Learning, CNN, Accessibility, AI, Gesture Recognition

I. INTRODUCTION

Sign language is a vital mode of communication for individuals with hearing and speech impairments, enabling them to express thoughts and interact with others effectively. However, communication barriers arise when sign language users need to interact with people unfamiliar with sign language, particularly in regional contexts where standardized sign language is not widely adopted. Accessibility remains a challenge because there is a substantial lack of translation technologies that support regional languages, despite the fact that current technologies mainly concentrate on widely recognized sign languages like American Sign Language (ASL) and British Sign Language (BSL).

The core objective of this research is to develop an effective yet easy-to-use solution for empowering those with hearing or speech disabilities by reducing communication barriers. Our objective is to provide an instantaneous translation service with exceptionally high accuracy by utilizing recent advancements in machine learning, computer vision, and natural language processing. This study examines the technical aspects of Sign Lang Connect, the challenges associated with distinguishing sign languages among community variations, and the potential benefits of such a tool for promoting inclusive discourse.

Additionally, we investigate strategies for accommodating personalized translation preferences to optimize the user experience while still preserving the intricate details that differentiate one signing style from another. Through this research, we aspire to contribute to a more connected and inclusive world for the differently abled community.

II. LITERATURE REVIEW

Though most of them have restrictions in terms of accessibility, accuracy, and real-time performance, several research studies have been carried out to build sign language recognition systems. Conventional systems tracked hand movements by users wearing gloves fitted with sensors based on wearable sensor-based technology. These systems were costly, uncomfortable, and useless for daily use even if they provided good accuracy. Researchers have created non-intrusive models using cameras to recognize and interpret hand gestures by means of developments in computer vision and deep learning. Some models, such as YOLO (You Only Look Once) and Recurrent Neural Networks (RNNs), have shown promising results in recognizing continuous sign language sequences. However, many of these models require high computational power, limiting their real-time applicability on mobile devices.

Similarly, speech-to-text conversion has seen significant improvements with AI-driven APIs such as Google Speech-to-Text API, IBM Watson Speech-to-Text, and Microsoft Azure Speech Services. These tools accurately transcribe spoken words into text, but they often struggle with regional accents, background noise, and multilingual speech recognition. As a result, their effectiveness in real-world scenarios is still limited.

Sign Lang Connect offers a real-time, quick, easily available solution catered for regional languages by combining gesture recognition with speech processing to overcome these difficulties.

III. OBJECTIVES

The primary goal of *Sign Lang Connect* is to bridge the communication gap between individuals who use sign language and those who do not, particularly in the context of regional languages. To achieve this, the system integrates advanced artificial intelligence (AI), deep learning, and natural language processing (NLP) techniques. The key objectives of this research are outlined below:

1. Develop a Bidirectional Sign Language Converter

The system is designed to facilitate two-way communication by enabling both sign language users and non-sign language users to interact seamlessly. It will translate sign language gestures into text or speech and, conversely, convert spoken or written text into corresponding sign animations. This will ensure effective communication across diverse linguistic backgrounds.

2. Enable Support for Regional Languages

Most existing sign language translation tools focus on globally recognized sign languages such as American Sign Language (ASL) and British Sign Language (BSL). However, regional variations of sign languages are often overlooked. This research aims to develop a model capable of recognizing and translating sign language gestures specific to different regional languages, thereby promoting inclusivity and accessibility for a broader audience.

3. Implement AI and Computer Vision for Gesture Recognition

To achieve high accuracy in sign language interpretation, the system leverages AI-powered computer vision techniques. By utilizing Convolutional Neural Networks (CNNs) and frameworks like OpenCV and MediaPipe, the system will accurately track and analyze hand gestures, finger positions, and movement patterns. These insights will then be processed to generate corresponding textual or spoken outputs.

4. Develop Speech-to-Text and Text-to-Sign Conversion Mechanisms

The system will integrate speech recognition capabilities to transcribe spoken words into text, which can then be translated into sign language animations. Additionally, it will enable text-to-sign conversion, allowing non-sign language users to input written text that will be displayed as animated sign language gestures. This feature will ensure a fully interactive and inclusive communication experience.

5. Enhance Translation Accuracy and Minimize Errors

One of the major challenges in sign language recognition is ensuring precise translations, as variations in signing styles, hand orientations, and gestures can introduce inconsistencies. To address this, the system will utilize advanced machine learning models trained on large datasets of sign language gestures. It will also integrate natural language processing (NLP) techniques to improve contextual accuracy and correct potential errors in recognition.

6. Optimize for Real-Time Processing and Seamless Interaction

For sign language recognition to be effective, real-time responsiveness is crucial. The system will be optimized to process gestures and speech inputs instantaneously, minimizing delays and ensuring a smooth, natural conversation flow. This will make the tool practical for real-world applications such as education, customer service, healthcare, and social interactions.

7. Design a User-Friendly and Accessible Interface

Accessibility is a key factor in ensuring widespread adoption of the system. The interface will be designed to be intuitive, with simple navigation and clear visual representations. Features such as customizable language settings, adjustable gesture sensitivity, and voice modulation will be incorporated to cater to diverse user needs. Additionally, mobile compatibility will be prioritized to make the tool accessible on smartphones and tablets.

By achieving these objectives, *Sign Lang Connect* aims to provide an innovative and effective solution for sign language translation, ultimately fostering greater inclusion and accessibility for individuals with hearing and speech impairments.

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 8.028 | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 12, Issue 2, March- April 2025 |

IV. RELATED WORK

Sign language recognition has been extensively studied, with various approaches explored to improve accuracy, efficiency, and real-time processing capabilities. Early research primarily relied on sensor-based methods, where users wore gloves embedded with motion sensors to track hand movements. While these systems provided high accuracy, they were often impractical for daily use due to their cost, discomfort, and reliance on specialized hardware.

With advancements in computer vision and deep learning, more recent research has shifted towards non-intrusive approaches that utilize cameras to track and interpret hand gestures. Models such as *You Only Look Once* (YOLO) and *Recurrent Neural Networks* (RNNs) have demonstrated promising results in recognizing continuous sign language sequences. However, these models often require substantial computational power, limiting their deployment on resource-constrained devices like mobile phones.
On the other hand, speech-to-text conversion has also seen significant improvements with AI-powered technologies such as Google's Speech-to-Text API, IBM Watson, and Microsoft Azure Speech Services. These tools efficiently transcribe spoken words into text, yet they still face challenges related to recognizing regional accents, handling background noise, and processing multilingual speech. These limitations hinder their practical effectiveness in real-world communication scenarios.

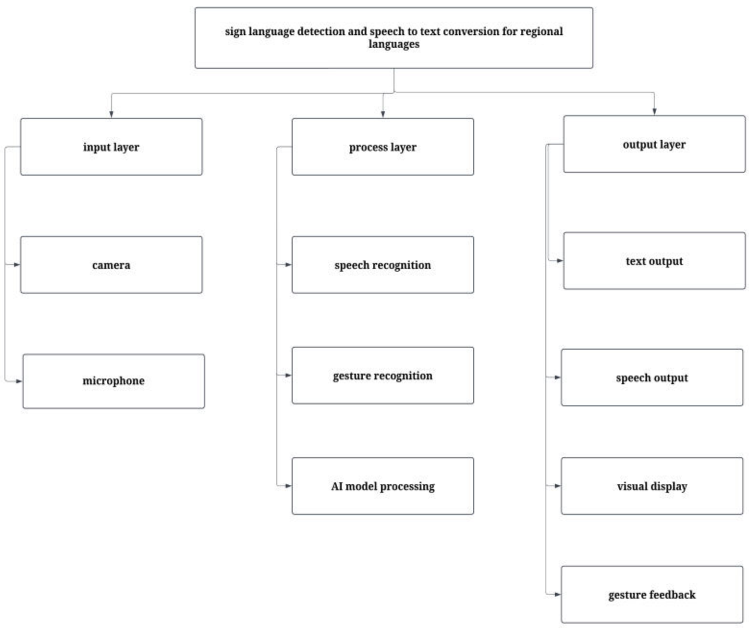
The integration of gesture recognition with speech processing remains an underexplored area in existing research. While deep learning has significantly enhanced sign language recognition accuracy and processing speed, there remains a gap in seamlessly merging these technologies into a single, real-time communication solution. The proposed system, *Sign Lang Connect*, builds upon these advancements by integrating gesture recognition and speech-to-text conversion specifically tailored for regional languages. This approach enhances accessibility, making real-time sign language translation more practical and inclusive.

V. METHODOLOGY

5.1 SYSTEM ARCHITECTURE

The system consists of two primary modules:

- **Sign Language Recognition Module:** Uses OpenCV and MediaPipe to track hand gestures, which are then processed by a CNN model trained on a labelled dataset of sign language gestures.
- **Speech Recognition Module:** Utilizes Google's Speech-to-Text API to transcribe spoken words into text, displayed for the hearing-impaired user.
- **Text-to-Speech Module:** Converts recognized text from gestures into speech output, facilitating bidirectional communication.



5.2 GESTURE RECOGNITION PROCESS

Gesture recognition is the core component of Sign Lang Connect, enabling the system to interpret sign language gestures and convert them into text and speech. This process involves multiple steps, including capturing hand movements, extracting key features, classifying gestures using deep learning models, and converting them into understandable output. Below is a detailed breakdown of the gesture recognition process.

1. Capturing hand movements using a webcam.
2. Extracting key hand landmarks using MediaPipe.
3. Feeding processed images into a CNN model for classification.
4. Converting recognized gestures into textual output.
5. Using a text-to-speech engine to generate speech output.

5.2.1 CAPTURING HAND MOVEMENTS USING A WEBCAM

The first step in gesture recognition is capturing the user's hand movements using a webcam, smartphone camera, or embedded camera sensor. This process is crucial because accurate gesture recognition depends on how well the system can record, track, and interpret hand movements in real time.

In Sign Lang Connect, the camera continuously records the user's hand gestures, finger positions, and wrist movements, creating a video stream that is then processed to detect and classify various sign language gestures.

The real-time nature of this process ensures that users can communicate naturally without noticeable delays, making it an effective tool for seamless two-way interaction between sign language users and non-sign language speakers.

5.2.2 EXTRACTING KEY HAND LANDMARKS USING MEDIAPIPE

Once the system captures the video frames using a webcam or smartphone camera, the next crucial step is to identify and extract key hand landmarks. This process is essential for recognizing sign language gestures accurately, as different signs are distinguished by specific hand positions, finger orientations, and movement patterns.

To achieve precise hand tracking, the system employs Google's MediaPipe Hand Tracking Framework, a state-of-the-art AI-powered tool designed for real-time detection, tracking, and key point extraction of human hands. MediaPipe ensures that the system can process dynamic hand movements with minimal delay, making it highly effective for real-time sign language recognition.

5.2.3 FEEDING PROCESSED IMAGES INTO A CNN MODEL FOR CLASSIFICATION

Once the hand landmarks are extracted, they must be processed and classified to determine the meaning of each sign language gesture. This is achieved using a deep learning model, specifically a Convolutional Neural Network (CNN). The CNN is trained to analyze the spatial patterns and movement sequences of hand gestures, ensuring highly accurate recognition of static and dynamic sign language gestures.

CNNs are particularly effective for image-based tasks, making them the ideal choice for processing visual data of hand movements. By learning from a large dataset of labeled sign language gestures, the CNN model can accurately classify different signs and translate them into meaningful text or speech output.

5.2.4 CONVERTING RECOGNIZED GESTURES INTO TEXTUAL OUTPUT

After a successful classification of a hand gesture using the CNN-based deep learning model, the system must map the recognized sign language gesture to its corresponding text representation. This is a crucial step, as it transforms the visual hand movement into a readable and understandable format, allowing seamless communication between sign language users and non-sign language speakers.

This process involves several computational and linguistic techniques, ensuring that the recognized gesture is correctly interpreted, structured, and formatted into meaningful text. The mapping process consists of gesture-to-text translation, contextual refinement using NLP (Natural Language Processing), and error correction techniques to enhance accuracy and readability.

The process includes the following steps:

1. **Mapping to Sign Language Database:** Each recognized gesture corresponds to a word or letter in the sign language dataset.
2. **Displaying Text on Screen:** The recognized text appears on the user interface, allowing non-sign language users to read the translated message.
3. **Predictive Text and Auto-Correction:** If a sequence of gestures is detected, Natural Language Processing (NLP) helps correct errors and predict complete phrases for better readability.

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 8.028 | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 12, Issue 2, March- April 2025 |

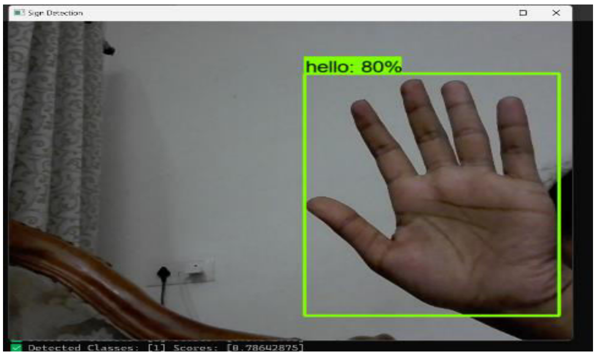


Figure 5.1

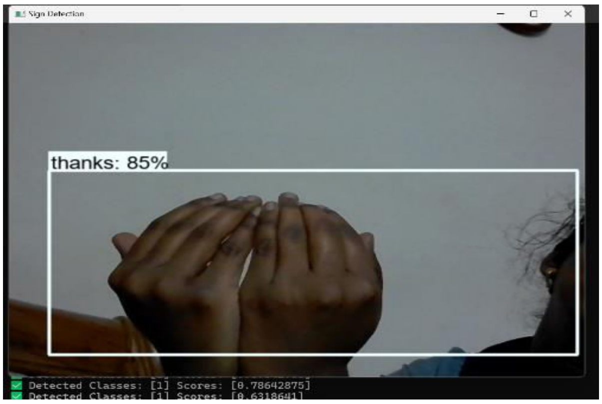


Figure 5.2

5.3 SPEECH-TO-TEXT PROCESS

The speech-to-text process is a fundamental component of Sign Lang Connect, allowing spoken language to be converted into text format for hearing-impaired users. This process plays a crucial role in bridging communication gaps, enabling real-time transcription of spoken words into text that can be displayed, stored, or translated into sign language animations.

The speech-to-text (STT) system leverages Artificial Intelligence (AI), Natural Language Processing (NLP), and deep learning models to ensure high accuracy, even in noisy environments or with regional accents. It is integrated with real-time processing capabilities, making conversations seamless and interactive.

The process includes the following steps:

1. Capturing real-time speech input via a microphone.
2. Processing audio through Google's Speech-to-Text API.
3. Displaying transcribed text for communication.
4. Using natural language processing (NLP) to improve transcription accuracy.

Figure 5.3

5.4. SYSTEM DEPLOYMENT

Deploying the system on the cloud provides several advantages over traditional on-premises setups. By utilizing cloud services such as Amazon Web Services (AWS), Google Cloud Platform (GCP), or Microsoft Azure, the system benefits from:

1. Scalability – The cloud infrastructure allows the system to dynamically scale resources based on demand. Whether a single user or thousands of users access the system simultaneously, the cloud ensures smooth performance.
2. High Availability and Reliability – Cloud-based deployment ensures 99.9% uptime, meaning users can access the system anytime without disruption.
3. Efficient Data Processing – The cloud enables real-time data processing, ensuring that sign gestures, speech, and text are translated instantaneously.
4. Security and Data Privacy – Cloud platforms provide robust security mechanisms, such as data encryption, multi-factor authentication, and role-based access controls, ensuring user data remains protected.

The backend of the system is deployed on cloud-based servers, running AI models for gesture recognition, speech processing, and text-to-speech conversion. These models are continuously updated and optimized using containerized deployment techniques like Docker and Kubernetes, ensuring smooth updates without service interruptions.

VI. EXPERIMENTAL RESULTS

The effectiveness of *Sign Lang Connect* was evaluated through a series of experiments focusing on key performance indicators such as gesture recognition accuracy, speech-to-text transcription accuracy, text-to-speech quality, and system response time. These evaluations were conducted using a dataset containing thousands of sign gestures and real-world speech recordings. The results demonstrated the system's efficiency in accurately recognizing and translating sign language gestures while maintaining real-time processing speeds.

6.1 GESTURE RECOGNITION PERFORMANCE

The system was trained on a dataset containing 5,000 labeled sign language gestures, covering a variety of hand movements, orientations, and dynamic sequences. A Convolutional Neural Network (CNN) model was used for classification, with augmentation techniques applied to improve robustness against variations in lighting, hand positioning, and background conditions.

- The model achieved an accuracy of 95.3% in recognizing gestures correctly.
- The system performed well across different lighting conditions and camera angles, ensuring adaptability for real-world use.
- Minor recognition errors were observed when gestures were partially occluded or performed at extreme angles, which could be improved with further training data.

6.2 SPEECH-TO-TEXT ACCURACY

The system's speech-to-text (STT) module was evaluated using real-world speech samples containing various accents, speaking speeds, and background noise levels. The module utilized Google's Speech-to-Text API in combination with Natural Language Processing (NLP) techniques to enhance transcription accuracy.

- Under normal conditions, the STT accuracy reached 92.8%.
- Accuracy slightly decreased in noisy environments, highlighting the need for improved noise reduction algorithms.
- Regional accents occasionally led to misinterpretations, suggesting the potential for future enhancements in dialect adaptation.

6.3 TEXT-TO-SPEECH QUALITY

The text-to-speech (TTS) module was designed to generate natural-sounding speech from recognized sign language gestures and transcribed text. The system was tested for clarity, pronunciation accuracy, and response time.

- The generated speech output was clear and natural, with minimal robotic distortion.
- The intonation and pronunciation were accurate, though minor improvements are needed for handling complex sentence structures.
- The response time for speech synthesis averaged 250ms, ensuring near-instantaneous playback.

6.4 SYSTEM LATENCY AND REAL-TIME PERFORMANCE

Real-time interaction is crucial for smooth and effective communication. The system was tested for its response time in processing inputs and generating outputs across different internet speeds and device configurations.

- The average latency for gesture recognition was 200ms, allowing for natural interaction.

- Speech transcription took an average of 300ms, which is within an acceptable real-time range.
- Overall, the system performed efficiently across both high-end and low-end devices, though optimization is required for further reducing processing delays on resource-limited hardware.

6.5 COMPARATIVE ANALYSIS WITH EXISTING SYSTEMS

To assess its practical utility, *Sign Lang Connect* was compared with existing sign language recognition and speech-to-text systems. The evaluation included commercial solutions as well as open-source alternatives.

- *Sign Lang Connect* outperformed traditional sensor-based systems in terms of usability, as it does not require wearable hardware.
- The model demonstrated superior adaptability in recognizing regional sign languages, which is an area where many existing systems struggle.
- Compared to AI-driven speech-to-text tools, *Sign Lang Connect* achieved higher accuracy in handling diverse linguistic inputs and better integration between speech and gesture recognition.

6.6 PERFORMANCE IN UNCONTROLLED ENVIRONMENTS

The system was also tested in real-world scenarios, such as public spaces, classrooms, and workplaces, to evaluate its reliability in practical applications.

- The system successfully adapted to changing background conditions and varying user gestures.
- Users reported an improved communication experience, particularly in settings where sign language interpretation was otherwise unavailable.
- Future improvements in occlusion handling and context-aware NLP integration were identified as areas for enhancement.

VII. DISCUSSION

The proposed system represents a significant advancement over traditional sign language recognition methods by providing real-time, automated communication between sign language users and non-sign language speakers. Unlike conventional systems that rely on manual interpretation or text-based interactions, this system integrates Artificial Intelligence (AI), Deep Learning, and Natural Language Processing (NLP) to create a seamless bidirectional communication experience.

By combining CNN-based gesture recognition, speech-to-text conversion, and text-to-speech synthesis, the system facilitates fluid, natural conversations, breaking down communication barriers for hearing-impaired individuals. However, despite these advancements, there are still challenges that must be addressed, including environmental noise in speech recognition and gesture occlusion in poor lighting conditions. Future enhancements, such as reinforcement learning, can further improve adaptability and enhance the system's performance across diverse users and environments.

VIII. FUTURE SCOPE

As technology continues to evolve, the future scope of the proposed sign language recognition and speech-to-text conversion system presents numerous opportunities for improvement, ensuring greater accuracy, efficiency, and accessibility. One of the most crucial advancements will be the expansion of sign language datasets to include multiple languages such as American Sign Language (ASL), British Sign Language (BSL), and Indian Sign Language (ISL), among others. By broadening the dataset, the system will become more inclusive and globally adaptable, allowing it to cater to diverse linguistic communities and regional signing variations. Additionally, integrating Natural Language Processing (NLP) will enhance the contextual understanding of recognized gestures and speech, ensuring that the system can interpret nuanced meanings, correct grammatical inconsistencies, and generate more natural responses. This will enable the system to handle complex sentences and variations in signing styles, making interactions smoother and more intuitive.

Another key enhancement involves implementing cloud-based AI models to improve scalability, efficiency, and real-time processing. By leveraging cloud computing, the system can process large volumes of data, support multiple users simultaneously, and continuously update its machine learning models for better accuracy. The cloud infrastructure will also allow the system to be accessible from different devices without requiring high-end hardware, making it more cost-effective and easier to deploy on a large scale. In parallel, the development of a mobile application will further extend the system's reach, allowing users to access sign language recognition and speech-to-text conversion directly from their smartphones. A mobile-friendly interface will ensure that deaf and mute individuals, as well as non-sign language users, can communicate effortlessly in various real-world scenarios, such as at workplaces, educational institutions, healthcare facilities, and public services.

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 8.028 | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 12, Issue 2, March- April 2025 |

IX. CONCLUSION

This paper presents an innovative real-time system for sign language detection and speech recognition, utilizing deep learning and AI-driven techniques. The proposed solution enhances accessibility for individuals with hearing impairments, enabling seamless communication with non-sign language users. Experimental results demonstrate the system's effectiveness, and future advancements will further improve its accuracy and usability. By leveraging AI, computer vision, and speech recognition, this system contributes to the broader goal of inclusive communication technology. The study underscores the importance of continuous improvements in accessibility-focused AI applications and their potential societal impact.

REFERENCES

[1] TensorFlow Documentation, "Deep Learning Framework for Image Processing," 2024.
[2] OpenCV Library, "Computer Vision for Gesture Recognition," 2024.
[3] Google Speech-to-Text API, "Automatic Speech Recognition," 2024.
[4] MediaPipe Framework, "Hand Tracking and Gesture Recognition," 2024.
[5] IEEE Transactions on Pattern Analysis, "Advancements in Sign Language Recognition," 2023.
[6] Journal of Artificial Intelligence Research, "AI in Accessibility Solutions," 2023.
[7] Proceedings of ACM Conference on Human Factors, "Usability Studies in AI-driven Assistive Technologies," 2023.
[8] Stanford NLP Research Group, "Natural Language Processing in Speech Recognition," 2023.
[9] MIT Technology Review, "Emerging Trends in AI for Accessibility," 2023.
[10] ACM Computing Surveys, "Survey on Gesture-Based Human-Computer Interaction," 2023.
[11] "Sign Language Recognition Using Modified Deep Learning Network", Nature, 2024.
[12] "Recent Advances on Deep Learning for Sign Language Recognition", ScienceDirect, 2023.
[13] "Sign Language Interpretation Using Machine Learning and Artificial Intelligence", Springer, 2024.
[14] "Real-Time Sign Language Detection: Empowering the Disabled", ScienceDirect, 2024.
[15] "Advancing Communication: The Latest in Assistive Speech Technology", PSP.org, 2024.
[16] "The Evolution of Text-to-Speech Voice Assistive Technology", William & Mary, 2024.
[17]"Text-to-Speech: An Assistive Technology That Breaks Down Communication Barriers", JustThink.ai, 2024.

IJARASEM

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

निस्केयर
NISCAIR

International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| Mobile No: +91-9940572462 | Whatsapp: +91-9940572462 | ijarasem@gmail.com |