iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70647

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Real-Time Sign Language Recognition in Digital Meeting

Shaik Mohammed Isaaq[1], Ganesh P[2], Rithvik[3], S Shashank Kumar[4], Dr. CN Sharif[5]
Dept of AIML, Ballari Institute of Technology and Management, Ballari, India.

Abstract: Real-Time Sign Language Recognition in Digital Meetings is an innovative solution designed to bridge the communication gap for individuals with hearing or speech impairments, fostering inclusivity and accessibility in digital communication. By integrating advanced technologies such as deep learning, WebRTC, and text-to-speech synthesis, this system enables the seamless conversion of sign language gestures into audible speech in real time, ensuring effective interaction between verbal and non-verbal participants. The solution leverages deep learning models to accurately recognize and interpret sign language gestures, while WebRTC facilitates low-latency, real-time communication, making it suitable for virtual meetings and online collaboration platforms. Furthermore, the integration of text-to-speech synthesis ensures that recognized gestures are transformed into clear, audible output, allowing for fluid and natural communication. This approach not only enhances accessibility for individuals with hearing or speech challenges but also creates an inclusive environment where all participants can engage and collaborate effortlessly. By addressing the communication barriers faced in digital meetings, this technology has the potential to revolutionize virtual interactions, making them more equitable and universally accessible for diverse users.

I. INTRODUCTION

Video conferencing has fundamentally transformed the way individuals and organizations communicate, becoming an indispensable medium for personal, professional, and educational interactions in today's increasingly digital world. The global shift to remote work and virtual engagements, particularly during the COVID-19 pandemic, catalyzed widespread adoption of platforms such as Zoom, Microsoft Teams, and Google Meet, which quickly evolved to incorporate advanced features like screen sharing, breakout rooms, live captions, AI-powered noise cancellation, and real-time transcription. These innovations have enhanced user experience, productivity, and collaboration; however, they have also highlighted a persistent gap in accessibility for individuals with hearing or speech impairments. Despite the rapid development of communication technologies, the lack of inclusive features for this group has posed significant challenges, restricting their ability to fully engage and participate in virtual meetings. Addressing this critical need, the Real Time Sign Language Recognition in Digital Meetings project introduces a groundbreaking solution designed to bridge this accessibility gap and promote inclusivity. By leveraging cutting-edge technologies such as WebRTC for low-latency communication, deep learning for accurate gesture recognition, and Text-to-Speech (TTS) synthesis for audible output, this project enables the seamless real-time conversion of sign language gestures into spoken audio. This not only allows individuals with hearing or speech impairments to communicate effectively with verbal participants but also fosters an environment where all users, regardless of their abilities, can interact, collaborate, and contribute on an equal footing. By transforming sign language into a universally understandable format in real time, this innovative system ensures that virtual communication platforms become equitable, empowering, and accessible to everyone, thereby redefining digital communication as a truly inclusive experience.

II. OBJECTIVES

A. Enable Seamless Real-Time Video and Audio Streaming Using WebRTC

The primary goal is to implement a communication framework that ensures uninterrupted, low-latency transmission of video and audio between participants. By integrating WebRTC (Web Real-Time Communication), the system facilitates high-quality peer-to-peer streaming, which is essential for capturing and transmitting sign language gestures in real time during digital meetings.

B. Develop a Machine Learning Model for Accurate Gesture Recognition

A robust machine learning model will be designed and trained to detect and recognize a wide range of sign language gestures. This model will process video frames captured from the user's camera, extracting meaningful features and classifying gestures with high precision. The focus is on achieving reliable recognition across different lighting conditions, hand movements, and signer variations to ensure practical usability.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

C. *Implement Text-to-Speech (TTS) Technology for Gesture-to-Audio Conversion*

To enhance communication between sign language users and verbal participants, the system will incorporate a Text-to-Speech engine. Once a gesture is recognized, it is converted into corresponding text, which is then synthesized into natural-sounding speech. This enables real-time audio feedback, making interactions seamless and inclusive for all participants in virtual meetings.

III. RELATED WORKS

A. *Literature Survey*

1) Deep Learning in Sign Language Recognition: A Hybrid Approach for the Recognition of Static and Dynamic Signs, the research tackles the problem of real-time sign language recognition using a hybrid deep learning model. The study addresses challenges such as varying speeds and durations of sign gestures, which make real-time detection difficult. It combines Long Short-Term Memory (LSTM) networks for dynamic gestures and YOLOv6 for static gesture recognition. The hybrid approach achieves high accuracy (96% for static signs and 92% for dynamic signs) by leveraging MediaPipe for feature extraction and integrating skeleton tracking. Despite these advancements, limitations remain in adapting the model for different signers and real-world variability, such as lighting and occlusion. The methodology shows promise for application in inclusive communication platforms like video conferencing, though further optimization for real world use is needed. A Conceptual Framework for Fee Automation System (2022) by Deepak Kumar Verma, Vishal Pandey, Deep Sagar Agrahari, Anubhav Rai: Focuses on a Java-based fee management system that integrates secure online payments, automated receipt generation, and real-time data storage to reduce errors, save time, and improve financial transparency during critical periods.
2) Real-time Sign Language Recognition using Machine Learning and Neural Networks, the focus is on integrating sign language recognition into practical applications, emphasizing the need for robust gesture interpretation in real-time. The study utilizes Media Pipe for feature extraction and experiments with various classifiers and neural networks, achieving an average accuracy of 97-98%. However, challenges include ensuring consistent frame capture during live video and reducing computational overhead. This work is particularly relevant to the proposed problem statement as it directly envisions extending sign language recognition to virtual communication platforms, though scalability and real time adaptability remain open challenges.
3) Sign Language Recognition System Using Convolutional Neural Network and Computer Vision, the research focuses on recognizing static American Sign Language (ASL) gestures using a CNN-based system and computer vision techniques. The methodology involves using HSV color space for background elimination, converting RGB images to grayscale, and extracting binary pixel features from resized 64x64 images for classification. A 2D CNN architecture with max pooling and SoftMax activation achieves over 90% accuracy for 10 ASL gestures, with low computational requirements making it suitable for real-time applications. Challenges include maintaining consistent hand positioning during dataset creation and handling gesture complexity. This system's approach aligns with real-time video conferencing needs, offering a robust and efficient foundation for enabling inclusive communication by translating sign language gestures into understandable formats.

IV. PROPOSED METHOD

The The proposed system aims to bridge the communication gap in virtual meetings for individuals with hearing or speech impairments by integrating real-time sign language recognition, gesture-to-speech conversion, and low-latency video communication.

1) To achieve real-time interaction, the system leverages WebRTC and WebSocket technologies to ensure seamless video and audio streaming with minimal latency.
2) The solution integrates a gesture recognition model and text-to-speech synthesis on the server side, allowing recognized signs to be instantly translated into audible speech, fostering inclusive digital communication.

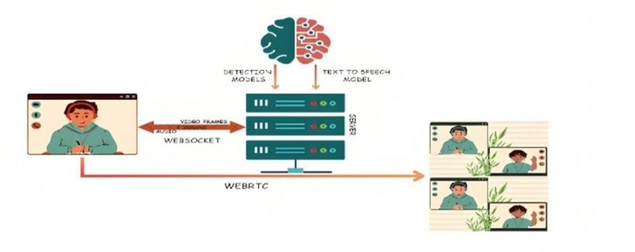


Figure 1 System Architecture

A. Architecture

The proposed methodology consists of multiple components that work together to achieve low-latency, gesture-aware communication during digital meetings. The key stages include:

1) Data Capture and Transmission: Video and audio streams are captured using the user's camera and microphone. These streams are transmitted to the backend server using WebSocket, enabling real-time, bidirectional data flow between the client and server.

2) Server-Side Processing

The backend server processes the incoming data using two major models:

- Detection Models: These models analyse video frames to identify hand gestures, facial expressions, or relevant movements that correspond to sign language.
- Text-to-Speech (TTS) Model: Recognized gestures are translated into textual form and then synthesized into speech for audio output.

3) Real-Time Streaming and Feedback: Processed video and audio outputs are sent back to users using WebRTC, ensuring high-quality peer-to-peer streaming. Real-time feedback like synthesized speech or visual overlays is also relayed for continuous interaction.

4) Workflow Loop and Automation: The entire process runs in a loop: initialization, frame capture, server-side processing, sign detection, and response transmission. The system notifies users of outcomes such as "No Sign Detected" when applicable.

5) Scalability and Optimization: The server architecture is optimized through clustering and load balancing to support multiple concurrent users. Efficient model deployment ensures balanced computational load and consistent performance under high demand.

B. Workflow

The flowchart illustrates the workflow of a real-time sign language gesture recognition system used during digital meetings, detailing the steps involved in capturing, detecting, and communicating sign language gestures effectively.

The process begins when a digital meeting is initiated, allowing users to join the platform to enable seamless real-time communication. As the meeting starts, the system automatically activates the user's camera and begins capturing image frames, which act as the primary input for gesture recognition.

The next step involves verifying whether a frame has been successfully captured. If the system fails to capture the frame, it attempts to retry, ensuring a continuous and uninterrupted data flow. Once a valid frame is captured, it is sent to the backend, where it undergoes processing using advanced machine learning models.

The core of the system involves analysing the captured frame using a trained prediction model to detect any sign language gestures being performed.

Based on the analysis, two possible outcomes are handled:

- If a gesture is detected, the system converts the interpreted sign into spoken audio using text-to-speech (TTS) technology. This audio is then forwarded to the other participants in the meeting, ensuring that the communication is effectively conveyed to all members.
- If no recognizable gesture is found in the frame, the system alerts the user by displaying a “No Sign Detected” message, allowing them to try again.

The system operates in a continuous loop, capturing and processing new frames in real-time to ensure that sign language communication remains fluid and responsive throughout the meeting session. The workflow concludes when the digital meeting ends, at which point the system halts the ongoing processes of frame capture, gesture detection, and audio communication.

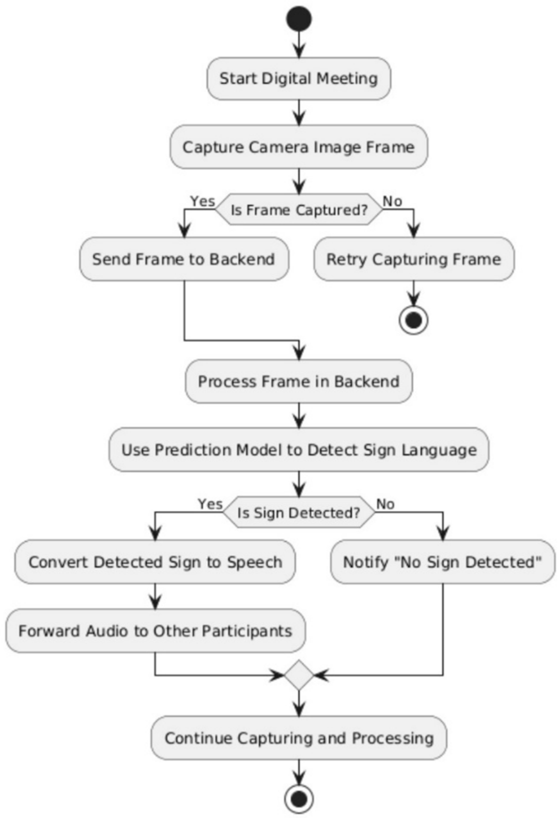


Figure 2 Data flow diagram

C. Sequence Models

The sequence diagram represents a system designed for real-time sign language detection and translation during a digital meeting, enabling seamless communication for users relying on sign language. The process begins when the user initiates the system, activating the camera to continuously capture video frames. The camera operates in a loop, ensuring frames are captured from the video feed. If a frame is successfully captured, it is sent to the backend for further processing. If the frame capture fails, the system automatically retries until it successfully acquires a frame. Once the backend receives the frame, it processes the data and forwards it to a prediction model, which analyses the frame to detect any sign language gestures.

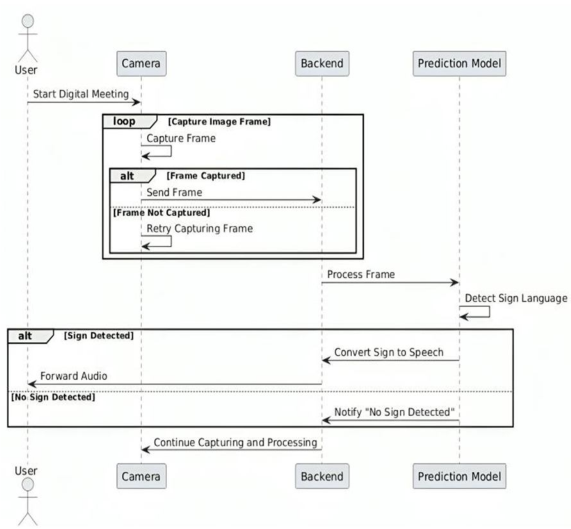


Figure 3 Student Sequence Models

D. UML (Unified Modeling Language)

In This use case diagram represents a system designed for real-time sign language detection and translation. Here's an explanation of its components: Actors:

- User: Represents the person initiating and interacting with the system.
- Prediction Model: A secondary actor responsible for detecting gestures from processed frames.

System Boundary:

- The rectangle labelled "System" encapsulates all the use cases handled by the system, showing the functions provided to the user.

Use Cases:

- Start Digital Meeting: The user initiates the system, starting the process of capturing frames.
- Capture Frame: The system captures video frames in real time.
- Process Frame: Captured frames are sent to the backend for processing.
- Detect Sign Language: The backend forwards the processed frames to the prediction model to identify sign language gestures.
- Convert Sign to Speech: If a sign is detected, it is converted to an audio message for the user.
- Notify No Sign Detected: If no sign language is detected, the system alerts the user.
- Forward Audio Output: Once the sign is converted to speech, it is forwarded to the user as audio output.

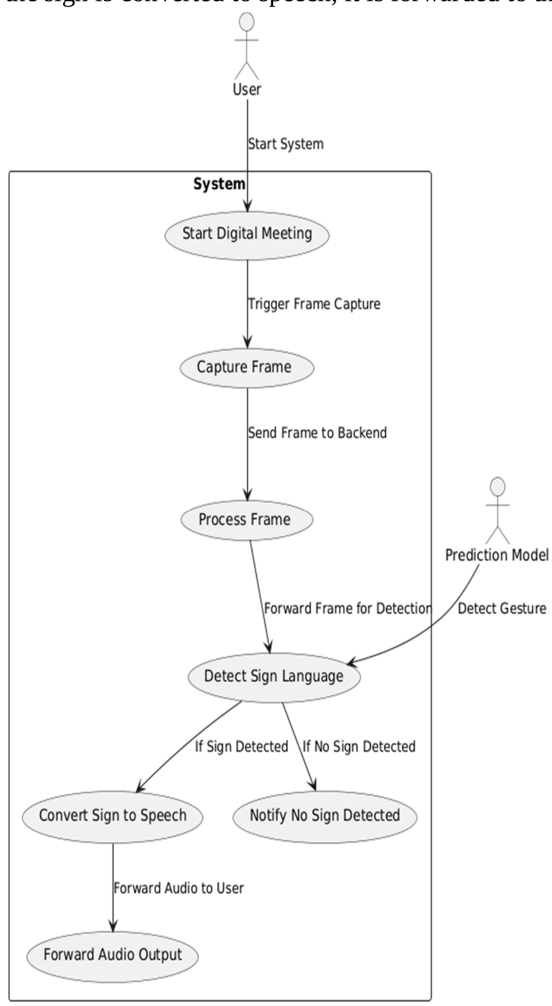


Figure 4 UML (Unified Modeling Language)

V. RESULTS AND DISCUSSION

The results of the test cases validate the functionality and efficiency of the sign language recognition system. The system performed well across various scenarios, ensuring that real time sign-to-speech conversion and gesture detection are accurately handled, even under challenging conditions such as blurry frames or high latency.

1) Detection of Rapid Hand Signs: Test Case 1 demonstrated that the system accurately detected and converted fast hand signs, confirming its capability to process dynamic hand movements without compromising the recognition quality.
2) Performance Under Latency Conditions: The system performed effectively even under weak backend server conditions Although there was a slight delay, the sign-to-speech conversion remained accurate, showcasing the system's robustness in less-than-ideal network conditions.
3) Handling No Camera Input: In Test Case 3, where the camera was disconnected, the system behaved as expected by ceasing processing, ensuring that the application did not attempt to process empty or incomplete input, avoiding unnecessary errors.

• **z** the model's recognition of the "Hello" sign and its ability to accurately convert it to speech were confirmed, validating the accuracy of the gesture recognition model for commonly used signs.

• Handling Unknown Gestures: In Test Case 5, the system appropriately skipped processing for unknown gestures, ensuring it only acted on valid input and providing a clear fallback for unrecognized signs.

• Clear Sign Detection: For clear and unambiguous signs (Test Case 6), the system demonstrated its effectiveness in accurate detection and conversion, ensuring high accuracy when the camera feed provided clear visual input.

• Handling Low-Quality or No Input: The system exhibited resilience in handling frames with no hand signs (Test Case 7) or blurry camera feeds (Test Case 8). Despite these challenges, the system continued processing frames, demonstrating robustness against variations in input quality.

```
Enter the names of the signs you want to collect data for:
Sign 1 name: hii
Sign 2 name: ok
Sign 3 name: thankyou
Sign 4 name: name
Sign 5 name: bye
```

Figure 5 Results

```
Classification Report:
              precision    recall  f1-score   support

         bye       1.00      1.00      1.00        38
         hii       1.00      1.00      1.00        33
        name       1.00      1.00      1.00        44
          ok       1.00      1.00      1.00        48
    thankyou       1.00      1.00      1.00        37

    accuracy                           1.00       200
   macro avg       1.00      1.00      1.00       200
weighted avg       1.00      1.00      1.00       200

Confusion Matrix:
[[38  0  0  0  0]
 [ 0 33  0  0  0]
 [ 0  0 44  0  0]
 [ 0  0  0 48  0]
 [ 0  0  0  0 37]]
```

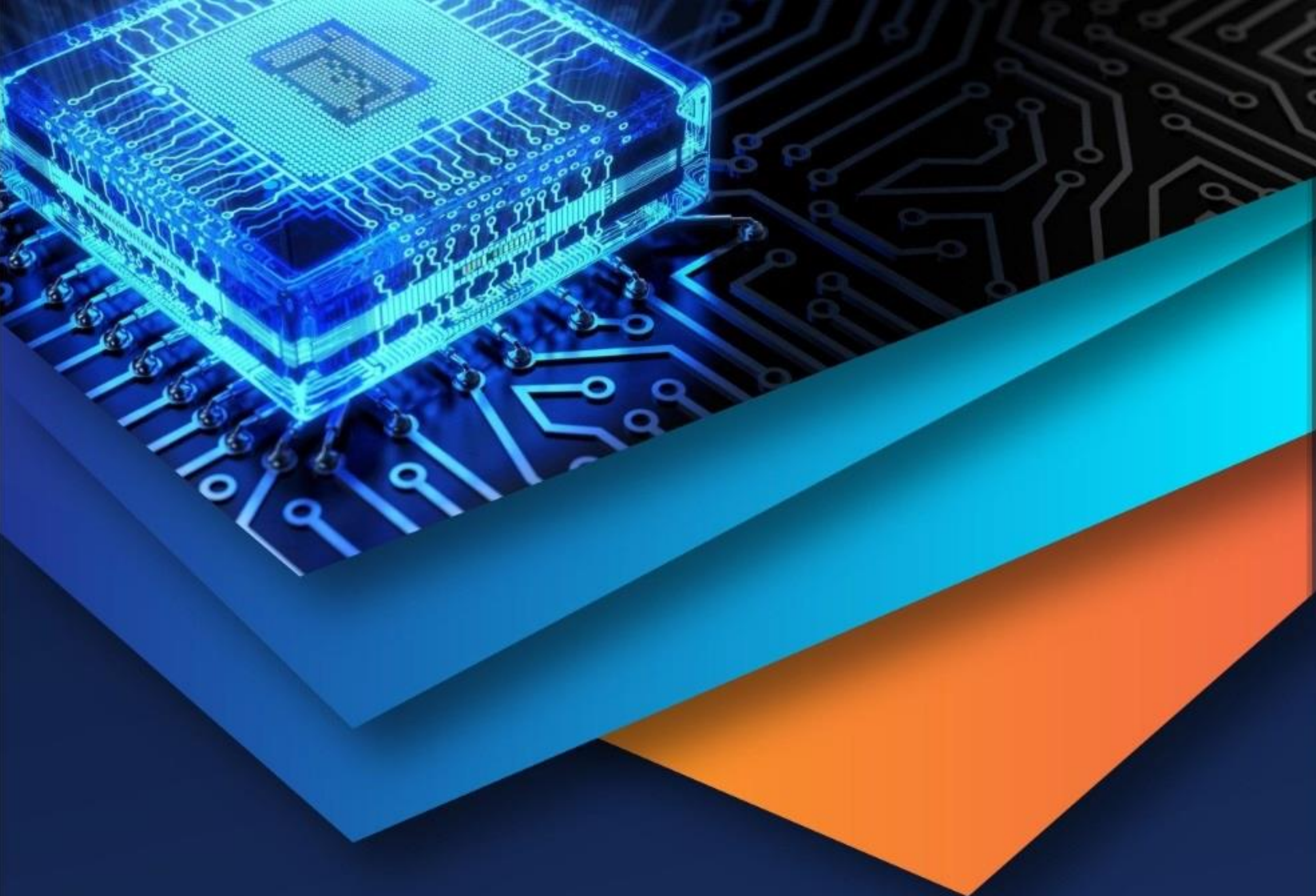
Figure 6 Results

VI. ACKNOWLEDGMENT

We extend our sincere gratitude to everyone who contributed to the successful completion of this project, *Online Fee Payment System.* We are also deeply grateful to the Head of the Department for their encouragement and valuable suggestions, which played a crucial role in shaping our work. We are also deeply grateful to the Head of the Department for their encouragement and valuable suggestions, which played a crucial role in shaping our work.

REFERENCES

[1] Romala Sri Lakshmi Murali, L.D. Ramayya, V. Anil Santosh: "Sign Language Recognition System Using Convolutional Neural Network and Computer Vision" December 2022, SSN (Online) 2582-1431 ISSN, PP 137-142
[2] Ahmed Mateen Buttar, Usama Ahmad, Abdu H. Gumaei, Adel Assiri, Muhammad Azeem Akbar, Bader Fahad Alkhamees: "Deep Learning in Sign Language Recognition: A Hybrid Approach for the Recognition of Static and Dynamic Signs" August 2023, DOI 10.17148/IJARCCE.2020.9607 SSN (Online) 2278-1021 ISSN.
[3] Roshnee Matlani, Roshan Dadlani, Sharv Dumbre, Shruti Mishra, Mrs. Abha Tewari: "Real-Time Sign Language Recognition Using Machine Learning and Neural Network" January 2022, ISSN 2582-7421.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY