iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60904

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

A Deep Learning Approach for Real-Time Sign Language Detection and Response through Avatar Model

Siva Sankari I[1], Shree Gayathri G[2], Sangeerane B[3], Devadharshini R[4], Dr. S. Sivakumari[5]
[1, 2, 3, 4]*UG Scholars,* [5]*Professor and Head, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, School of Engineering, Coimbatore, India*

Abstract: ***This paper focuses on the evolution of a real-time sign language detection model using computer vision, machine learning, and deep learning. Its goal is to narrow the poor communication for the speaking and hearing-impaired community by using recurrent neural networks and long short-term memory models. The proposed system utilizes a dataset comprising sign language gestures captured in various contexts and by different signers. Preprocessing techniques are applied to draw out fitting features from the video frames, including hand movements, facial expressions, and body postures. The LSTM neural network architecture is chosen to grab temporal dependencies in sequential data, making it suitable for the robust nature of sign language. The training process involves optimizing the LSTM network on the labeled dataset, incorporating techniques such as transfer learning and data augmentation to enhance model generalization. The resulting model is capable of recognizing a divergent set of sign language actions in real-time.***
Keywords: Preprocessing, Transfer learning, Data augmentation.

I. INTRODUCTION

The real-time sign language detection system intends to create an innovative system using computer vision and machine learning models to observe and identify sign language gestures in real time. This project points to the interaction gap between individuals with hearing impairments who speak sign language and those who cannot comprehend it. By leveraging advanced algorithms and models, the system tracks hand movements, extracts features, and employs neural networks to accurately classify gestures. The ultimate goal is to enable instantaneous translation of sign language into either contextual or verbal form, aiding competent communication between sign language speakers and the broader community in various real-world settings.

II. RELATED WORK

Mali et al. [1] (2023) presented "Sign Language Recognition Using Long Short-Term Memory Deep Learning Model" aiming to facilitate interaction for speech and hearing-impaired individuals. The study focused on employing neural network models, particularly MediaPipe Holistic and the Long Short-Term Memory model, for recognizing sign language gestures. MediaPipe Holistic, known for its precision and low latency in capturing pose, face, and hand keypoints, was integrated with LSTM to achieve accurate sign language recognition. The research specifically targeted commonly used words in American Standard Sign Language (ASL), achieving a notable accuracy rate of 98.50%.

Mhatre et al. [2] (2022) addressed the communication challenges faced by speech and hearing-impaired individuals by developing a real-time sign language detection system using deep learning. Their computation, using a Long Short-Term Memory (LSTM) model, achieved a high training accuracy of 90-96%. The system recognized seven commonly used Marathi sign language gestures, providing output in text and audio formats through the Google Text-to-Speech library. This innovation aims to empower individuals with speech and hearing impairments by facilitating easier and more accessible communication through technology.

Aparna et al. [3] (2020) proposed an innovative method for sign-language recognition using deep learning, employing Convolutional Neural Network, and one of the RNN models is the Long Short-Term Memory model. The CNN served as a pre-trained model for feature extraction, with the extracted features progressed to LSTM for capturing spatio-temporal information. To enhance accuracy, an additional LSTM was stacked. The study focused on the Indian Sign Language dataset evaluation, showcasing the performance of their algorithm. They emphasized the scarcity of research in deep learning architectures, particularly CNN and LSTM models which are used for sign language detection. Highlighting the importance of capturing temporal information in deep learning models, they underscored the need for further exploration in this domain.

©IJRASET: All Rights are Reserved

Sundar et al. [4] (2022) explored American Sign Language (ASL) alphabet recognition using MediaPipe and LSTM, leveraging advances in artificial intelligence. Gesture recognition, crucial for various applications including communication for the deaf-mute, human-computer interaction, and medical fields, was the focus. They devised a vision-based approach utilizing MediaPipe to capture hand landmarks, combined with a custom dataset for experimentation. LSTM was employed for hand gesture recognition, achieving an impressive 99% accuracy across 26 ASL alphabets.

Deshpande et al. [5] (2022) introduced a deep learning model for real-time Indian Sign Language used for gesture recognition and translation into words and content. Given the prevalence of ISL in the hearing-impaired group across India, the development of ISL recognition systems is crucial. Unlike other sign languages, ISL primarily features two-handed signs, adding complexity. Their system, utilizing a convolutional neural network which is considered Deep performs both feature extraction and the classification, pre-existed with the image pre-processing. Operating on present input taken from a webcam, the system delivers output in words and speech formats.

The built CNN architecture achieved a splendid 98% correctness across a dataset consisting 56 items, including digits, letters, and common words, offering a valuable communication tool for individuals with disabilities.

Pathak et al. [6] (2022) introduced a real-time sign language detection system, aiming to boost interaction between the deaf and the public. Their model, based on a Convolutional Neural Network (CNN), employed Transfer Learning with a Pre-Trained SSD model MobileNet V2 model architecture on a custom dataset.

The robust model consistently classified sign language gestures with high accuracy, offering significant benefits to sign language learners for practice. The study explored various human-computer interface procedures for posture recognition, ultimately favoring image processing techniques with human movement classification. Despite challenges like uncontrolled backgrounds and varying lighting conditions, the system achieved a commendable accuracy of 70-80% in recognizing selected sign language signs, marking a notable advancement in real-time sign language detection technology.

Kothadiya et al. [7] (2022) addressed communication barriers for individuals with speaking or hearing imparities by proposing "DeepSign," which is a deep learning model for sign language detection and recognition. Leveraging LSTM and GRU models, which are feedback-based learning models, the system detects and recognizes words from gestures in Indian Sign Language which is used for video frames. Different fusions of LSTM and GRU layers were explored, with the created model achieving an impressive correctness of around 97% over 11 different signs using their dataset, IISL2020. This original approach offers the potential to promote communication between those unfamiliar with sign language and individuals with speech or hearing impairments, thereby reducing communication barriers in society.

Kodandaram et al. [8] (2021) explored Sign Language Recognition (SLR) by exercising Deep Learning, a vital tool for improving communication accessibility for the deaf and speech-impaired. They highlighted the complexity of recognizing both static and dynamic hand gestures and proposed leveraging Convolutional Neural Network architectures for this task. Through training epochs, the model learns to identify hand gestures, generating corresponding English text and converting it to speech. This streamlined approach aims to enhance communication efficiency, making communication easier for individuals with hearing or speech impairments. The study underscores the importance of deep learning techniques in advancing Sign Language Recognition systems.

Sharma and Singh [9] (2021) developed a Deep Learning Model which is made for Indian Sign Language (ISL) Recognition, making three primary contributions.

They curated a large ISL dataset from 65 users in diverse settings and augmented it to enhance intra-class variance. Their new Convolutional Neural Network (CNN) architecture effectively extracted features and classified ISL gestures. Evaluation on multiple datasets yielded accuracies of 92.43%, 88.01%, and 99.52%. Additionally, the model exhibited efficiency in terms of processing time and achieved promising results compared to existing methods.

Deep et al. [10] (2022) proposed a Real-time Sign Language Detection and Recognition system utilizing advanced technologies. Initially, Regions of Interest (ROI) are found and tracked through skin segmentation using OpenCV. Subsequently, MediaPipe [15] captures hand landmarks, storing key points in a NumPy array and then the model is trained with the help of packages such as TensorFlow, Keras, and LSTM. The system enables real-time testing with live webcam feeds. This innovation holds promise for the deaf and mute community, facilitating their connection with the world. Unlike previous methods reliant on machine learning algorithms trained on images, this method leverages deep learning models, and reinforces real-time sign detection and recognition capabilities.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

III. METHODOLOGY

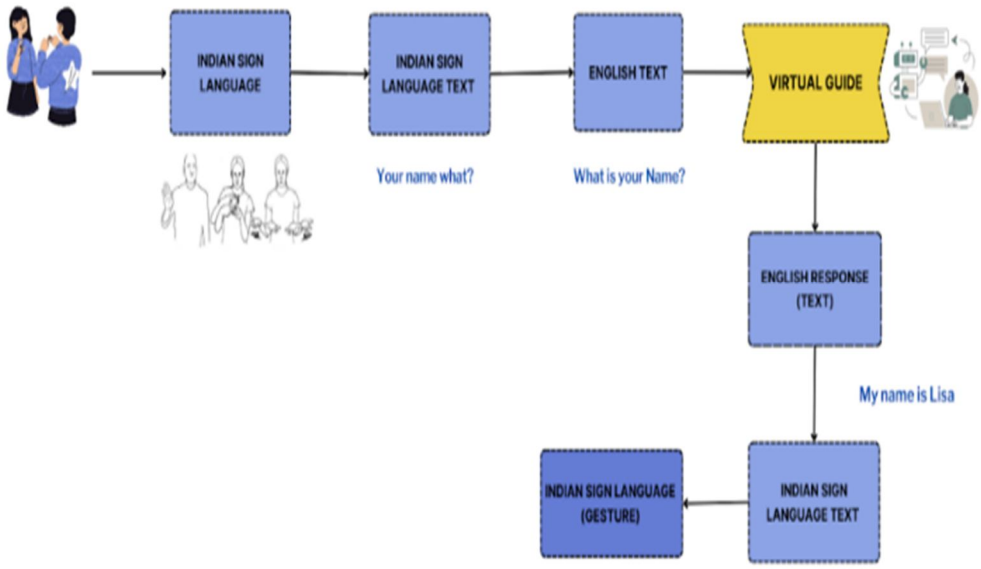


Fig 1: Block Diagram

IV. SYSTEM IMPLEMENTATION

1) *Indian Sign-Language Gesture Detection:* The first step is to recognize and trace the hand and body movements in Indian Sign Language gestures. This could be achieved by extracting relevant features [11] from the video or image frames, such as hand shape, joint positions, or motion trajectories.
2) *Feature Extraction:* Once the gestures are detected, features are extracted from the tracked hand and body movements. These features could include spatial information, temporal dynamics, or any other relevant characteristics that capture the essence of the sign language gestures.
3) *Sequence Modeling with LSTM:* The extracted features are then fed into a sequence model, such as a Long Short-Term Memory network which is a type of RNN. LSTM models are suitable for handling sequential data and can effectively capture temporal dependencies in the gestures.
4) *Indian Sign Language to Text Conversion:* The LSTM model processes the sequence of features and learns to map them to corresponding Indian Sign Language text representations. It learns the mapping between the gesture sequences and words of Indian Sign Language.
5) *Conversion to English Text:* Once the Indian Sign Language text is obtained, it can be further processed using translation techniques to convert it into English text.
6) *Generating Response in English:* The English text obtained from the translation step is then used to generate a response (from the virtual guide)
7) *Response to Indian Sign Language Text:* To respond in Indian Sign Language, the generated English text needs to be converted back to Indian Sign Language text. This can be done by using a reverse translation approach using the LSTM model.
8) *Feature Extraction from Videos:* Once the Indian Sign Language text is obtained, the corresponding video clips for the individual words can be retrieved from the dataset (by utilizing a large dictionary of isolated Indian Sign Language words mapped to their corresponding videos). Features are then extracted from these videos, similar to the initial gesture detection step, to capture the visual information necessary for generating the 3D avatar.
9) *3D Avatar Generation:* The extracted features from the videos are used to generate a 3D avatar that represents the sign language gestures. This involves animating the avatar's movements to match the captured features, creating a visual representation of the Indian Sign Language gestures.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

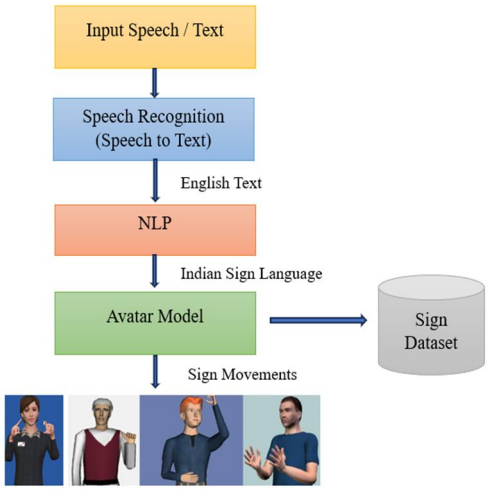


Fig 2: System Implementation Diagram

V. EMPLOYMENT OF LSTM IN SIGN LANGUAGE DETECTION

Long Short-Term Memory (LSTM) networks play a crucial role in capturing the temporal dependencies [13] inherent in the sequence of features extracted from sign language gestures. LSTMs are one of a kind of recurrent neural network (RNN) designed to overpower the disadvantages of oldest RNNs when dealing with longer-range dependencies.

On the subject of this Indian Sign Language (ISL) gesture detection system:

1) *Sequential Input Processing:* The extracted features, representing hand and body movements, are inherently sequential [12] in nature, forming a time series. LSTMs are well-suited for processing such sequential data.
2) *Memory Retention and Update:* LSTMs excel at capturing long-term dependencies by maintaining a memory cell. This cell retains information over extended periods, allowing the model to remember relevant context from earlier frames [14] as it processes subsequent ones. This is crucial for understanding the evolving dynamics of sign language gestures
3) *Effective Handling of Variable-Length Sequences:* Sign language gestures can vary in duration and complexity. LSTMs accommodate variable-length sequences, dynamically adjusting their internal state based on the input data. This flexibility is crucial for accurately modeling the diverse range of sign language expressions.
4) *Learning Temporal Patterns:* LSTMs inherently learn temporal patterns and relationships within the sequential input. This is essential for understanding the nuanced structure of sign language gestures, where the order and timing of movements convey meaning.
5) *Mapping to Text Representation:* The LSTM model processes the sequential features and learns to map them to corresponding Indian Sign Language text representations. By capturing temporal dependencies, the model discerns the linguistic nuances embedded in the sequence, enabling accurate conversion to textual representations.

VI. CONCLUSION

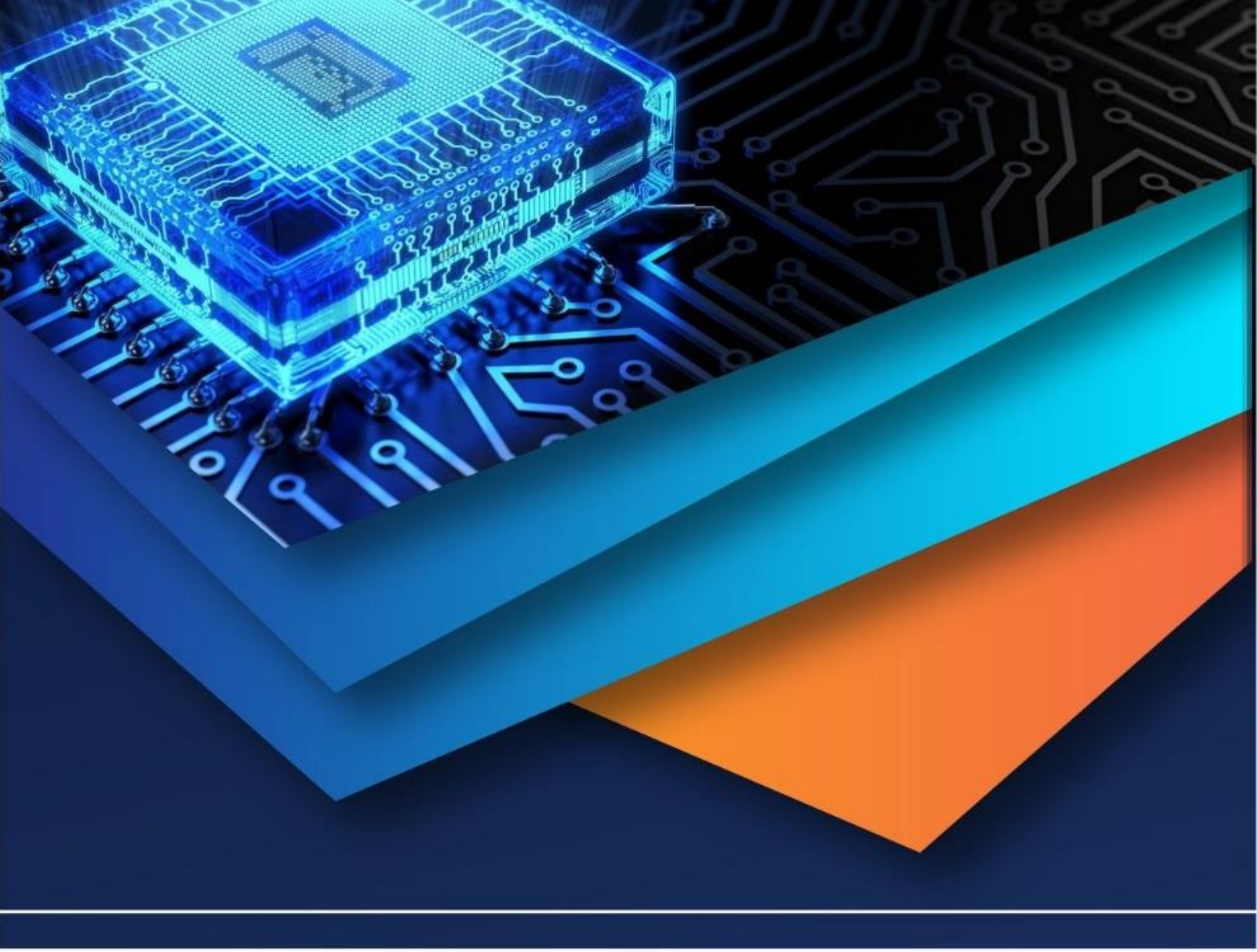
In the realm of real-time sign language detection, the past year has witnessed significant strides toward a more inclusive and accessible communication landscape. Leveraging the power of cutting-edge technologies, particularly in computer vision and deep learning, has enabled the evolution of systems capable of accurately interpreting sign language gestures at once. This progress marks a pivotal step forward, breaking down barriers for the hearing and speaking-impaired community and fostering seamless communication. However, amidst the celebration of these achievements, it is imperative to acknowledge persistent challenges that demand attention in future research endeavors.

REFERENCES

[1] Prabhat Mali, Aman Shakya & Sanjeeb Prasad Panday, “Sign Language Recognition Using Long Short-Term Memory Deep Learning Model”, ICIPCN 2023: Fourth International Conference on Image Processing and Capsule Networks pp 697–709.
[2] Shreyas Mhatre; Sarang Joshi; Hrushikesh B. Kulkarni, “Sign Language Recognition Using Long Short-Term Memory Deep Learning Model”, 2022 IEEE International Conference on Current Development in Engineering and Technology (CCET).

[3] C. Aparna & M. Geetha, “CNN and Stacked LSTM Model for Indian Sign Language Recognition”, Part of the book series: Communications in Computer and Information Science ((CCIS,volume 1203)).
[4] B Sundar, T Bagyammal, “American Sign Language Recognition for Alphabets Using MediaPipe and LSTM”, 4th International Conference on Innovative Data Communication Technology and Application, 2023.
[5] Ashwini M. Deshpande, Gayatri Inamdar, Riddhi Kankaria & Siddhi Katage, “A Deep Learning Framework for Real-Time Indian Sign Language Gesture Recognition and Translation to Text and Audio”, 2022
[6] Aman Pathak, Avinash Kumar, Priyam, Priyanshu Gupta, Gunjan Chugh, “Real Time Sign Language Detection”, International Journal for Modern Trends in Science and Technology 2022
[7] Deep Kothadiya, Chintan Bhatt, Krenil Sapariya , Kevin Patel , Ana-Belén Gil-González and Juan M. Corchado, “ Deepsign: Sign Language Detection and Recognition Using Deep Learning”, Electronics 2022, MDPI
[8] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.
[9] Sakshi Sharma & Sukhwinder Singh et al, “Recognition of Indian Sign Language (ISL) Using Deep Learning Model”, Volume 123, pages 671–692, 28 September 2021
[10] Aakash Deep, Aashutosh Litoriya, Akshay Ingole, Vaibhav Asare, Shubham M Bhole, Shantanu Pathak et al, “Realtime Sign Language Detection and Recognition”, 2022 2nd Asian Conference on Innovation in Technology (ASIANCON) IEEE.
[11] Bayegizova A, Murzabekova G, Ismailova A et al (2022) “Effectiveness of the use of algorithms and methods of artificial technologies for sign language recognition for people with disabilities”, Eastern-Euro J Enterprise Technol.
[12] Soodtoetong, N., Gedkhaw, E.: The efficiency of sign language recognition using 3D convolutional neural networks. In: 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). IEEE (2018)
[13] Papastratis I, Chatzikonstantinou C, Konstantinidis D, Dimitropoulos K, Daras P, “Artificial Intelligence Technologies for Sign Language Sensors “, 2021,21, 5843.
[14] Adaloglou N, Chatzis T, “A Comprehensive Study on Deep Learning-based Methods for Sign Language Recognition”, IEEE Trans Multimed, 2022,24, 1750–1762.
[15] Sharma S & Singh, “Vision-based sign language recognition system: A Comprehensive Review”, IEEE International Conference on Inventive Computation Technologies (ICICT), 2020.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)