iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68336

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 IV Apr 2025- Available at www.ijraset.com

Automated Sign Language Interpretation

Prof. S. A. Shikalgar[1], Jadhav Shivaji Rajendra[2], Kale Vishwajeet Vijay[3], Khedkar Utkarsh Sudhir[4], Ghutukade Swapnil Dada[5], Katawate Mayuresh Sandesh[6]
[1]Professor [2, 3, 4, 5, 6]Students, Department of Computer Science and Engineering, AGTI's Dr. Daulatrao Aher College of Engineering, Karad

Abstract: Communication plays an essential role in human interaction, allowing individuals to express ideas and emotions. While spoken languages are widely used, individuals with hearing and speech impairments rely on sign language. However, the lack of widespread understanding of sign language creates communication barriers between them and the hearing community. This study presents a real-time Indian Sign Language (ISL) recognition system using the Media pipe framework and Long Short-Term Memory (LSTM) networks. The approach involves training an LSTM model to distinguish between different signs, utilizing a dataset generated through a pre-trained Holistic model from the Mediapipe framework, which serves as a feature extractor.
Keywords: Sign Language, Mediapipe, LSTM, Computer Vision

I. INTRODUCTION

Effective communication is crucial for social interaction, but it is challenging for individuals who are deaf or mute. Due to the limited understanding of sign language among the general population, these individuals face difficulties in integrating into society. Sign language recognition (SLR) technology plays a significant role in assistive communication, allowing those with hearing impairments to interact seamlessly. Without such advancements, they may struggle with daily interactions, accessing services, and participating in social events. By developing reliable SLR systems, we can improve accessibility, independence, and inclusion for the hearing-impaired community. This study explores the application of LSTM networks in recognizing ISL. The primary objective is to detect, classify, and translate ISL gestures accurately using neural networks and computer vision techniques. ISL differs significantly from American Sign Language (ASL), making recognition complex. The challenges in ISL recognition include intricate hand movements, variations in individual signing styles, and background distractions. However, LSTM networks have shown potential in enhancing recognition accuracy and efficiency.

II. LITERATURE SURVEY

[1] The authors proposed a machine learning-based method for interpreting Indian Sign Language using contour-based feature extraction and classification. Their approach works effectively for static gestures, offering a low-cost solution suitable for basic applications. However, the method lacks support for dynamic or continuous gestures, limiting real-world usability.

[2] This study focused on developing a real-time ISL recognition system using CNNs integrated with OpenCV. The system achieved good results for static gestures with live camera input. Despite its promising real-time capability, the model was limited to a small vocabulary and could not handle continuous gesture streams.

[3] In this paper, handcrafted features such as contour orientation and Hu moments were used with a neural network for ISL recognition. The approach is computationally light and interpretable, performing well for isolated gestures. However, it falls short in recognizing complex or dynamic gestures, and the absence of temporal modeling restricts continuous recognition.

[4] A decade-long systematic literature review identified major trends and challenges in sign language recognition, including signer dependency, dataset scarcity, and limited support for continuous recognition. The review emphasized the growing role of deep learning and called for more robust, real-time, and signer-independent systems.

[5] A transformer-based architecture was introduced to model spatial and temporal patterns in sign language recognition. The approach demonstrated superior performance over traditional CNN-RNN models, especially for continuous gesture sequences. It highlights the potential of attention mechanisms for enhancing temporal understanding.

[6] This paper provided an extensive survey of Vision Transformers (ViTs), outlining their effectiveness in visual tasks compared to CNNs. Although not specific to sign language, the insights support the use of transformer models like [5], encouraging further exploration of ViTs for complex gesture understanding.

[7] The authors reviewed deep learning approaches in sign language recognition, comparing CNNs, RNNs, and hybrid models. They discussed key challenges like signer variability, lack of large datasets, and real-time processing constraints. Their analysis provides a strong foundation for choosing suitable deep learning architectures in SLR tasks.

III. CONCEPTS AND THEORIES BEHIND ASLI

A. *Computer Vision*

Computer vision enables the capture and processing of visual data from video inputs, identifying and analyzing hand gestures. This involves feature extraction techniques focusing on hand shape, movement, and orientation to facilitate accurate classification.

B. *Recurrent Neural Network*

RNNs have a feedback mechanism that allows information to be passed from one step of the sequence to the next as shown in Figure 1. The vanishing gradient problem is a significant disadvantage of classical RNNs, where the gradients used to update the network's weights become very small, making it difficult for the network to learn long-term dependencies. To address this issue, many RNN variations, such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), have been created, which contain additional methods to limit the propagation of information through the network and prevent the vanishing gradient problem.

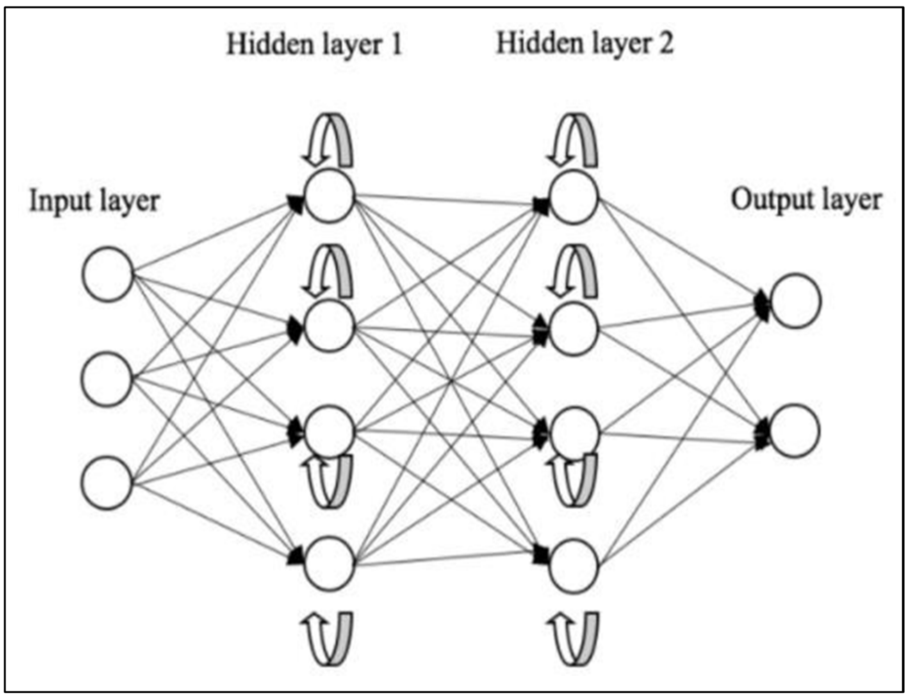

Fig 1 : Recurrent Neural Network

C. *Long Short Term Memory Network*

LSTMs have an elaborate design that incorporates "memory cells" and "gates" that govern the movement of data across the network, as opposed to traditional RNNs, that utilize a basic feedback loop to send information gained from one time step to a subsequent one. An LSTM receives an input vector as well as a hidden state vector containing data from the preceding time step at each time step. The input and hidden state vectors are then processed by the network through a set of gates that govern the information that flows through and out of the memory cells. The gates are composed of sigmoid functions which return values that span 0 to 1, indicating which elements of the input information as well as hidden state vectors must be permitted into the memory cells. Memory cells store data over multiple time steps, enabling the network to detect long-term dependencies in input data. The LSTM output is a combination of the current memory cell state as well as the hidden state vector at each time step, and it has the potential to be used for prediction or classification.

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 IV Apr 2025- Available at www.ijraset.com

IV. IMPLEMENTATION

The implementation work-flow for Automated Sign Language Interpretation.

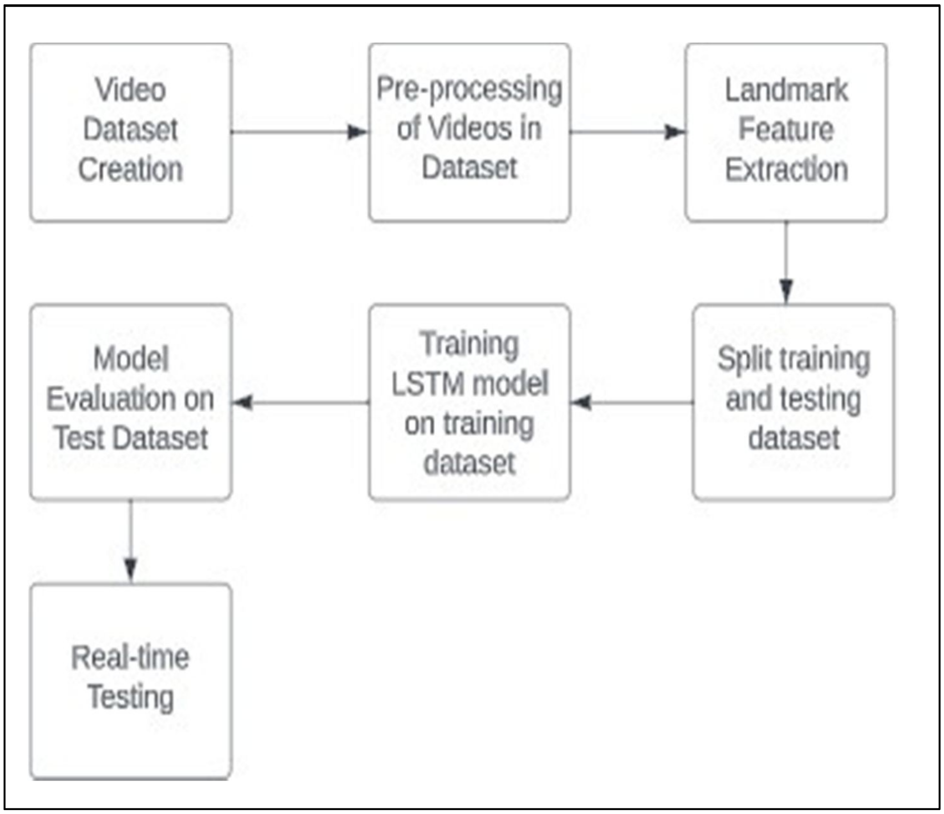

Fig 2 : Block Diagram of Dynamic ISL Gesture Recognition

A. *Dataset Creation and Pre-Processing*

Each dynamic sign language gesture was captured on film 30 times for the dataset utilized in this investigation. Each video is made up of 15 frames which was necessary in order completely film the sign gesture. In signs which didn't need all 15 frames, they were augmented to include zeros so that it reached 15 frames.

B. *Feature Extraction*

For capturing the required features that make up the dynamic sign gesture, certain landmarks were taken into consideration. The Mediapipe framework was employed to pick the right and left hand, face, and pose markers for this operation. Since dynamic gestures include more than just hand movement, the Holistic model was utilized which includes all three models to determine the co-ordinates of landmarks on hand, pose and face.

C. *Training the Model*

Split the preprocessed dataset into training and validation sets. Train the model on the training set using the extracted features.

D. *Model Evaluation*

Assess the model's performance with evaluation metrics, using the validation set to determine its accuracy, precision, recall, and F1-score. The hyper-parameters and network architecture were adjusted to optimize performance

E. *Real-Time Testing*

The trained model was later deployed on real-time scenarios to evaluate its performance recognizing Dynamic Sign Gestures.

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 IV Apr 2025- Available at www.ijraset.com

V. RESULTS AND ANALYSIS

The SLR system for dynamic gestures was able to achieve a training accuracy of 95.5% within 250 epochs. The model was trained employing different batch sizes, starting from the default 32, 64 and 128. The model with batch size of 128 performed better in comparison to the others.

After 250 epochs the model's accuracy started dropping with increase in loss. This indicates that the model reached its optimal performance after 250 epochs and any further training did not yield significant improvements. The 128 batch size proved to be effective in achieving high accuracy and efficiency. The model achieved an accuracy of 95.3% after being tested on the validation dataset. The confusion matrix of each dynamic gesture is as depicted The precision, recall and f1- score of each dynamic sign gesture is calculated. Furthermore the model was also tested with real-time sign gestures using a standard laptop camera where the gestures were done with no sign gesture isolation. The model was able to recognize the signs accurately but when switching from one to next, the response led to few false positives since the camera was picking up each change and trying to recognize the gesture. This suggests that exists a future scope to carry out further research with respect to successive dynamic sign recognition.

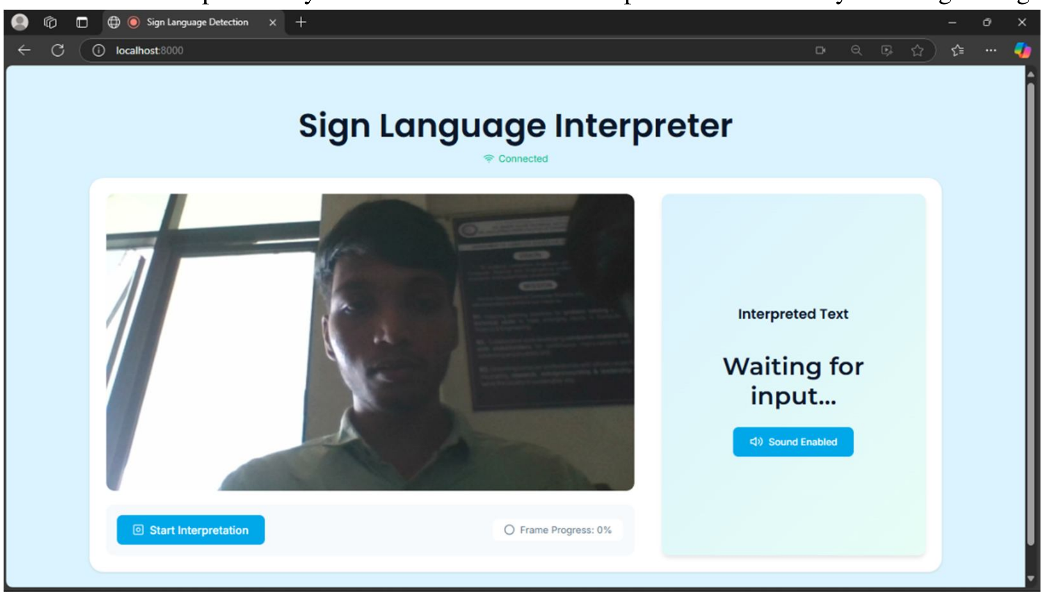

Fig 3 : Web based execution of project

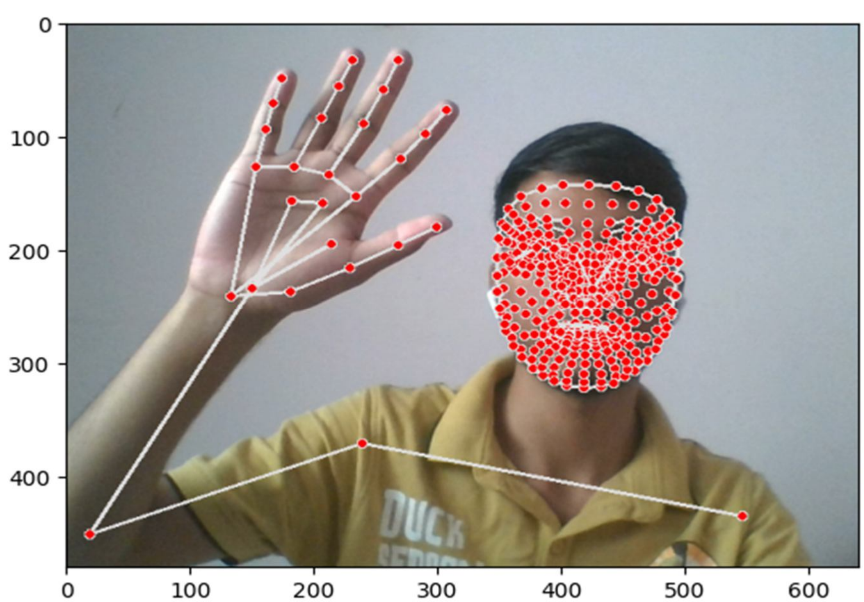

Fig 4 : Testing in real-time – Landmark Detection

VI. CONCLUSION AND FUTURE WORK

In conclusion, the study successfully developed and evaluated a SLR system for dynamic gestures in ISL recognition. Further research can focus on exploring different architectures and hyper-parameters to potentially improve accuracy even further.

Additionally, investigating the use of SLR systems for other sign languages and expanding the dataset could yield valuable insights and advancements in the field of gesture recognition. Furthermore, it would be interesting to investigate the impact of incorporating temporal information into the SLR system for dynamic gestures. This could involve exploring recurrent neural network architectures or attention mechanisms to capture the sequential nature of sign language.

Moreover, conducting user studies to evaluate the usability and effectiveness of the SLR system in real- world scenarios would provide valuable feedback for improving its practical applications.

Overall, the findings from this study lay the foundation for further studies in the domain of ISL recognition and pave pathways for the creation of more robust and accurate gesture recognition systems.

REFERENCES

[1] D. S, K. H. K B, A. M, S. M, D. S and K. V, "An Efficient Approach for Interpretation of Indian Sign Language using Machine Learning," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 2021, pp. 130-133, doi: 10.1109/ICSPC51351.2021.9451692.

[2] K. Shenoy, T. Dastane, V. Rao and D. Vyavaharkar, "Real-time Indian Sign Language (ISL) Recognition," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bengaluru, India, 2018, pp. 1-9, doi: 10.1109/ICCCNT.2018.8493808.

[3] P. C. Badhe and V. Kulkarni, "Artificial Neural Network based Indian Sign Language Recognition using hand crafted features," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225294.

[4] A. Wadhwan, and P. Kumar, “Sign Language Recognition Systems: A Decade Systematic Literature Review”, Archives of Computational Methods in Engineering, Springer, 2019, DOI: https://doi.org/10.1007/s11831-019-09384-2

[5] D. R. Kothadiya, C. M. Bhatt, T. Saba, A. Rehman and S. A. Bahaj, "SIGNFORMER: DeepVision Transformer for Sign Language Recognition," in IEEE Access, vol. 11, pp. 4730-4739, 2023, doi: 10.1109/ACCESS.2022.3231130.

[6] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, ‘‘A survey on vision transformer,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 1, pp. 87–110, Jan. 2023.

[7] M. Al-Qurishi, T. Khalid and R. Souissi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," in IEEE Access, vol. 9, pp. 126917-126951, 2021, doi: 10.1109/ACCESS.2021.3110912.