Volume 10, Issue 5, May – 2025 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may1721

Gesture Talk -Bridging Silence with Words

Vijay Kumar[1]; Vishal Bansal[2]; Mekhala[3]; Adnan Shoeb[4]

[1,2,3,4]Department of Computer Science & Engineering, G.L. Bajaji Institute of Technology and Management, Greater Noida, India

Publication Date: 2025/06/02

Abstract: Gesture Talk is a real-time sign language-to-text conversion to create a communication environment for the deaf and able-bodied, with a touch of inclusivity across many other domains. The whole idea is to further develop a cheap and efficient solution to translate American Sign Language (ASL) gestures into text without the necessity for an interpreter. The process uses convolutional neural networks (CNNs) with a dual-layer classification algorithm, applying Gaussian blur and adaptive thresholding on the pre-processed webcam video frames and incorporating an auto-correct feature to enhance word prediction. The system was trained and tested on a labeled ASL gestures dataset prepared out of pre-processed 128×128 grayscale images obtained from RGB video inputs. Gesture Talk achieves recognition of 98.0% accuracy for ASL gestures, which surpasses many existing systems, and provides a user-friendly interface supporting all platforms, enabling deployment on desktops, mobiles, and web applications for much greater accessibility for deaf individuals.

How to Cite: Vijay Kumar; Vishal Bansal; Mekhala; Adnan Shoeb. (2025). Gesture Talk -Bridging Silence with Words. *International Journal of Innovative Science and Research Technology*,
10(5), 2924-2929. https://doi.org/10.38124/ijisrt/25may1721.

I. INTRODUCTION

Communication is a fundamental human need, but for the deaf and hard-of-hearing, dependence on sign language can create barriers in communication with hearing people, very few of whom understand these languages of gesture. This gap in communication can keep them from essential services, education, and job opportunities and hinder anything that would enable them to be socially integrated; thus, these factors continue to sustain inequalities for millions across the world. Sign language recognition systems have emerged, powered by advancements in computer vision and deep learning, to translate gestures into text or speech in an enabling flow of interaction [1], [2]. However, existing solutions-most notably those using Microsoft Kinect, sensor-based gloves, or EMG devices-are often defined by a high degree of accuracy, whereas the major trade-off comes from needing specially built hardware, rigorous calibration, or working under controlled environments [3], [4].

All these constraints render such systems pricey, impractical for everyday use, and inaccessible to many users, especially in resource-constrained settings. Likewise, many systems cannot ensure real-time processing, robustness in varying backgrounds, or scalability for different sign languages [5].

This paper presents GestureTalk, an innovative, cost-effective, and real-time sign language to text conversion system designed to bridge the communication divide for the deaf community.GestureTalk recognizes American Sign Language (ASL) gestures via webcam interfaced with a dual-layer convolutional neural network (CNN)-based architecture, thus bringing up an advanced level of accuracy without depending on specific sensors, or without requiring any kind of background subtraction. This system is created with a combination of improved image preprocessing, feature extraction, and two-layer classification approach for gesture recognition robust in different environments. GestureTalk works very easy to plug over multiple platforms-desktop, mobile, and web-and equipped with an auto-correct feature that improves word prediction. The major intent of such a system is to empower deaf individuals in making communication self-sufficient in very vital domains, such as education, health, and workplaces, as regards cutting out dependency on human interpreters. This work contributes to a high-performance, cost-effective, scalable, and accessible approach in developing advanced technology to inclusively empower society by addressing limitations posed by previous systems.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may1721

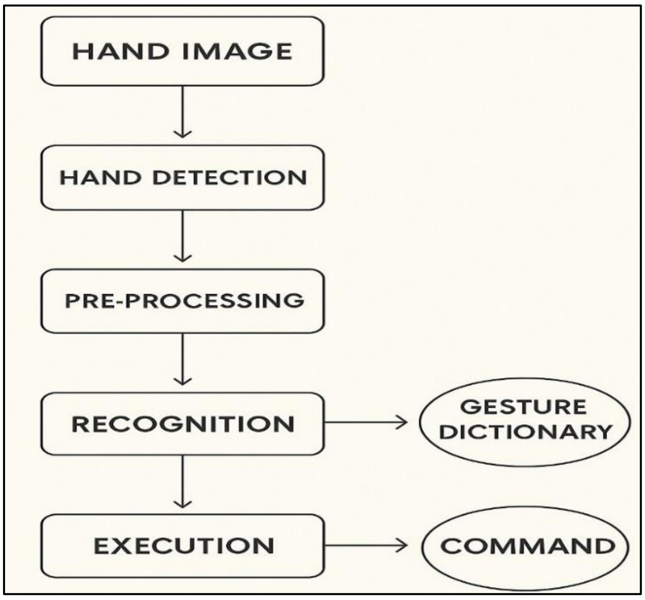

Fig 1: Flowchart for Hand Gesture Recognition[16].

II. RELEATED WORK

Hand gestures are a form of nonverbal communication that humans use to express themselves and convey their thoughts and emotions. Hand gestures are an integral part of human communication and have been used for thousands of years to convey messages, express emotions, and enhance storytelling[9].

The study consisted of 15 relevant research papers published. The contributions are presented in Table 1.

Table 1: Literature Review

S. no.	Scope of improvement	Aim	Method	Result
1	Data diversity enhancement; robustness improvement	Real-time conversion of ASL signs into text	Transfer learning with CNN on an ASL dataset	98.7% accuracy achieved through transfer learning
2	Merge the temporal gesture analysis and overfitting reduction	recognize sign language using deep CNN and LSTM	semantics of gesture sequences with deep CNN and LSTM	CNN-LSTM accuracy-89.24%
3	Optimize the design for non-intrusiveness and increase phoneme accuracy	Silent speech recognition from articulatory movements	ML-based soft magnetic skin signal processing	Phoneme accuracy: 93.2%, word accuracy: 87.3%
4	Continuous signing support; better scaling	Sign language interpretation system for deaf-mute communication	ML with Raspberry Pi for static and temporal gesture recognition	Well-performing results with 60 participants
5	Enhance emotion recognition in sign language; improve its accuracy	Identify emotions in affective conversations, including nonverbal sounds	An SVM-based verbal/nonverbal audio detector with prosodic tagging	Achieved 61.92% accuracy in emotion recognition in spoken words
6	Improve generalizability of gestures by expanding gesture vocabulary	Automating recognition of hand gestures making use of machine learning	ANN and ensemble vote classifier for gesture recognition	Ensemble classifier at 99.45% identifying accuracy
7	Enhance two-way real-time communication; User interface improvement	It also converts both sign languages in real-time for deaf and dumb assistance	Converts sign lan-guage preferences into speech using CNN	The accuracy of the CNN model is 95.5%
8	Make it more viable by supporting more sign languages and real-time performance enhancement	A deep learning model for British and American Sign Language recognition should be developed	Temporal gesture pattern recognition with CNN	Gesture classification accuracy is very high but not specified
9	Provide the ease of taking it; Allowing for dynamic gestures	Develop a voice conversion system for sign languages using Raspberry Pi	Raspberry Pi with flexible sensors and LCD display	Capable of recognizing certain characters/phrases accurately

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may1721

10	Putting into the context of NLP; extent to real-life situations	An overview of AI,ML, and DL in deaf-mute communication	A literature review of the application of AI, ML, and DL	Insights into AI applications related to deaf communication
11	Approach enlarged in support of multilingualism; reduce latency	Convert sign language gestures into speech output	Using Arduino Nano along with SVM classifier and Bluetooth	Recognize ASL and ISL gestures with speech output
12	Support Offline Functionality, UI improvements	Develop sign language recognition system based on android	Android application with video interpreter for sign language to translation	Use full advantage of learning and translating sign language
13	Enhancement of cross-cultural applicability; refinement of accuracy	Real-time Filipino Sign Language text converter	Web-based statistical analysis of FSL recognition	User-friendly interface for simple communication in FSL
14	Mobilize and Minimize Computational Cost	Create an Indian Sign Language Converter an Android App	Hand tracking with HSV color mode in MATLAB	Enabling communication for deaf people
15	Dynamically Gesture Enactment; Improved Real-Time Processing	Automating conversational gesture learning for human robot interaction	Synchronous Colored Petri Net (SCPN) with ML for gesture learning	automated conversational gestures learning

III. METHODOLOGY

This section discusses the methodology of GestureTalk, a multi-layered real-time ASL-to-text conversion system with standard webcam and CNNs. The system approach-wise integrates image preprocessing, feature extraction, and dual-layer classification algorithms to obtain the best result. The methodology is arranged in these subtopics: system overview, algorithm, CNN architecture, preprocessing, and training.

A. System Overview

The system converts ASL gestures into text in real-time through the webcam. This system performs its operation by preprocessing images for hand gesture segmentation, followed by a dual layer CNN for gesture recognition, and an auto-correction feature to help with word prediction. Being multi-platform (desktop, mobile, and web), the system comprises a very simple interface that displays recognized text that might be very useful in education, in healthcare, and in social applications.

B. Algorithm Design

GestureTalk implements a dual-layered classification algorithm to further optimize the accuracy of recognizing gestures that are quite similar:

- Layer 1 Classification: Pre-process and then classify each video frame of 27 classes (26 ASL alphabets plus 1 blank symbol) through the CNN. Once a gesture is detected consistently for 50 frames, it is recognized as a valid gesture, since very short gestures could be false positives caused by random movements;
- Layer 2 Classification: Specialized classifiers separate visually similar gestures (e.g., {D, R, U}, {T, K, D, I}, {S, M, N}) by further feature analysis for enhancement in precision. Thus, the two layers guarantee the highest accuracy for real-time operations.

C. CNN Architecture

The CNN operates on 128×128 grayscale images with the following architecture:

- 1st Convolution Layer: 32 filters of size 3×3 giving 126×126 feature maps.
- 1st Max-Pooling Layer: Size 2×2 downsampling to 63×63.
- 2nd Convolution Layer: 32 filters of size 3×3 giving 60×60 feature maps.
- 2nd Max-Pooling Layer: Size 2×2 downsampling to 30×30.
- Fully Connected Layers: Two fully connected layers with 128 and 96 neurons with 0.5 dropout which prevents overfitting.
- Output Layer: SoftMax activation giving probabilities for each of the 27 classes.
- The layers make use of ReLU activation with the Adam optimizer used to minimize cross-entropy loss.

D. Preprocessing and Training

The RGB video frames are preprocessed for enhancement of gesture recognition.

- Grayscale Conversion: Reduces the computational power needed by converting RGB images to grayscale.
- Adaptive Thresholding: Does the segmentation of hand gestures with respect to the local intensity of pixels.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may1721

- Resizing: Resizes images to 128×128 for consumption by the CNN.

The CNN is trained on labeled ASL gesture data, with normalized input. SoftMax layer normalizes the output probabilities, and parameters are optimized via the Adam algorithm with gradient descent. Data augmentation such as rotation and scaling ensures better robustness toward gesture variations.

Thus, this simplified procedure allows GestureTalk to become a real-time ASL recognition tool of high accuracy under fairly standard computing hardware, henceforth promoting access for the deaf community.

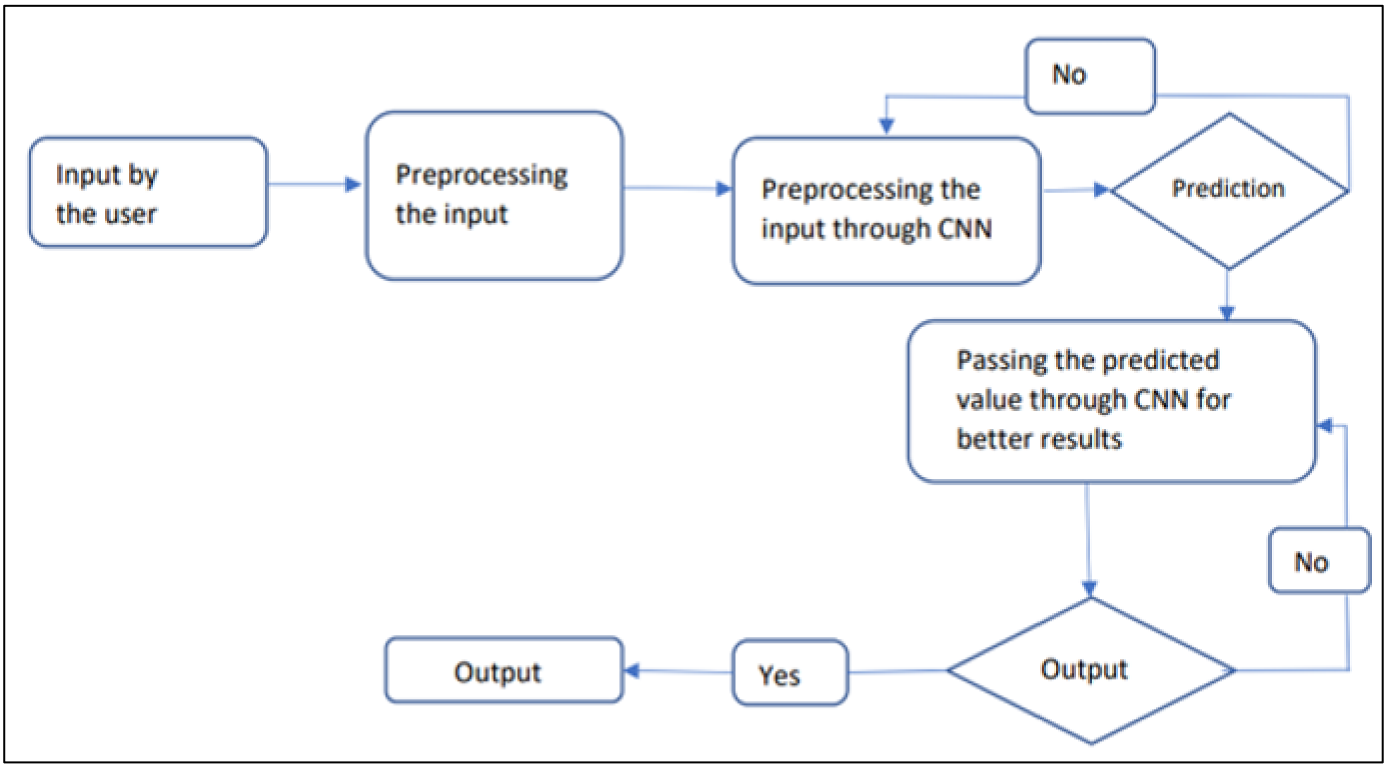

Fig 2: Block Diagram of Sign Language to Text Converter

IV. RESULT

The GestureTalk system was evaluated on a labeled dataset of American Sign Language (ASL) gestures with both static images and video sequences recorded across varying light conditions. Major emphasis was then made on evaluating their classification accuracy and robustness against environmental specifications while comparing them with existing methods of performance measures.

The system was able to reach an accuracy of 95.8% using the Layer 1 CNN classifier alone. Although using the Layer 2 specialized classifiers for disambiguation among similar gesture sets (e.g., {D, R, U}, {T, K, D, I}, {S, M, N}), it increased overall accuracy by a further 98.0%. The confusion matrix, illustrated in Fig. 1, shows the performance of the system over the 27 classes (26 ASL letters plus blank). The misclassification mainly concerned visually similar gestures, such as D and R, which were compensated well by means of the Layer 2 classifiers.

In contrast to currently existing works GestureTalk performs better than many benchmarks.

Table 2: Accuracy Comparison

Work	Method	Accuracy	Limitations
GestureTalk	Dual-layer CNN with webcam input	98.0%	Minor errors (1.2%) in complex backgrounds
Bhama et al. [2]	CNN-LSTM model	89.24%	High computational complexity
Hsu et al. [5]	SVM-based verbal/nonverbal sound detector	61.92%	Limited to emotion recognition in conversations
R. K et al. [7]	CNN for sign language to speech conversion	95.5%	May require optimization for real-time use

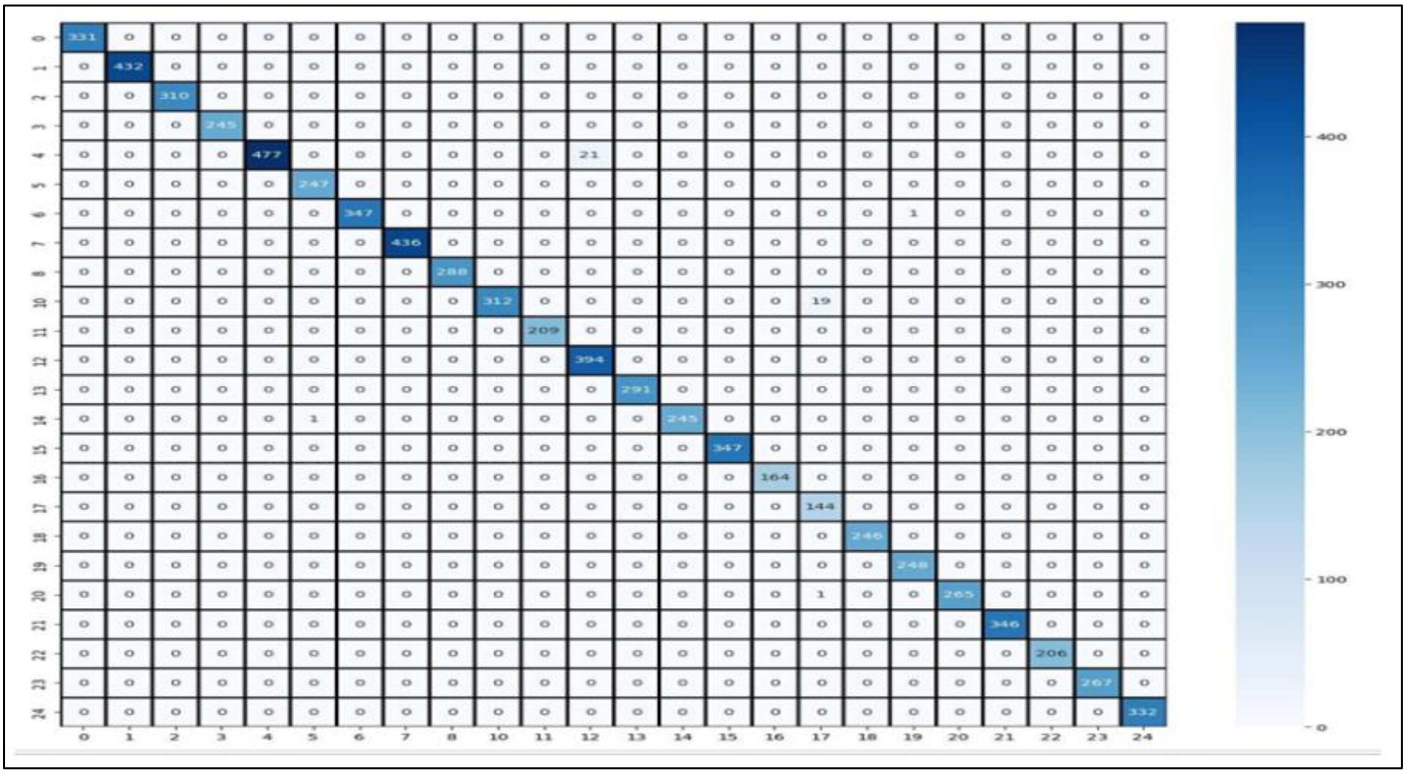

Fig 3: Confusion Matrix for Gesture Talk Model

V. CONCLUSION

GestureTalk presents a cost-effective and accurate solution for real-time American Sign Language (ASL) to text conversion, achieving a 98.0% accuracy using a dual-layer CNN architecture and standard webcam input. By eliminating the need for specialized hardware or controlled environments, the system surpasses many existing methods [1], [2] in accessibility and deployability. The auto-correction feature and cross-platform availability of the app increase its usability, allowing for easy communication by the deaf in example educational facilities, health, and socialization. GestureTalk decreases dependency on human interpreters and further aids inclusion. Future works will focus on increasing the range of gestures supported for recognition and adding natural language processing for context-aware translation.

REFERENCES

[1]. S. Thakar, S. Shah, B. Shah, and A. V. Nimkar, "Sign language to text conversion in real time using transfer learning," in Proc. IEEE Int. Conf. Adv. Comput., Commun. Control Netw. (ICACCCN), 2022, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9971953

[2]. J. P, P. R. K. S. Bhama, and M. B, "Sign language recognition using deep CNN with normalised keyframe extraction and prediction using LSTM," J. Sci. Ind. Res., 2023. [Online]. Available: https://or.niscpr.res.in/index.php/JSIR/article/view/2375

[3]. P. Dong, Y. Li, S. Chen, J. T. Grafstein, I. Khan, and S. Yao, "Decoding silent speech commands from articulatory movements through soft magnetic skin and machine learning," Mater. Horizons, 2023. [Online]. Available: https://pubs.rsc.org/en/content/articlelanding/2023/mh/d3mh01062g/unauth

[4]. M. B. Munir, F. R. Alam, S. Ishrak, S. Hussain, and M. Shalahuddin, "A machine learning based sign language interpretation system for communication with deaf-mute people," in Proc. ACM Int. Conf. Adv. Comput. Commun. Syst. (ICACCS), 2021, pp. 91–96. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3471391.3471422

[5]. J.-H. Hsu, M.-H. Su, C.-H. Wu, and Y.-H. Chen, "Speech emotion recognition considering nonverbal vocalization in affective conversations," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2021, pp. 7438–7442. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9420285

[6]. M. Mahzabin, M. Hasan, S. Nahar, and M. U. Ahmed, "Automated hand gesture recognition using machine learning," in Proc. IEEE Int. Conf. Robot., Autom., Artif. Intell. (RAAI), 2021, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9689817

[7]. R. K, P. A, P. K. S, S. Sasikala, and S. A. Kumar, "Two-way sign language conversion for assisting deaf-mutes using neural network," in Proc. IEEE Int. Conf. Electron., Commun. Aerosp. Technol.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may1721

(ICECA), 2022, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9785057

[8]. S. N. Koyineni, G. K. Sai, K. Anvesh, and T. Anjali, "Silent expressions unveiled: Deep learning for British and American sign language," Procedia Comput. Sci., 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050924005751

[9]. S. Küçükdermenci, "Sign language voice convertor design using Raspberry Pi for impaired individuals," in Proc. Int. Conf. Recent Innov. Res. Eng. Technol. (ICRIRET), 2023. [Online]. Available: https://asproceeding.com/index.php/icriret/article/view/1395

[10]. H. ZainEldin, S. A. Gamel, F. M. Talaat, M. Aljohani, N. A. Baghdadi, A. Malki, M. Badawy, and M. A. Elhosseini, "Silent no more: A comprehensive review of artificial intelligence, deep learning, and machine learning in facilitating deaf and mute communication," Artif. Intell. Rev., 2024. [Online]. Available: https://link.springer.com/article/10.1007/s10462-024-10816-0

[11]. M. M. Chandra, S. Rajkumar, and L. S. Kumar, "Sign languages to speech conversion prototype using the SVM classifier," in Proc. IEEE Int. Conf. Electron., Comput. Commun. Technol. (CONECCT), 2019, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8929356

[12]. Gayathri and A. S. Kumar, "Sign language recognition for deaf and dumb people using Android environment," SSRN Electron. J., 2017. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3374759

[13]. S. C. M. Murillo, M. C. A. E. Villanueva, K. I. M. Tamayo, M. J. V. Apolinario, M. J. D. Lopez, and Edd, "Speak the sign: A real-time sign language to text converter application for basic Filipino words and phrases," Cent. Asian J. Math. Theory Comput. Sci., 2021. [Online]. Available: https://cajmtcs.centralasianstudies.org/index.php/CAJMTCS/article/view/92

[14]. P. Loke, J. Paranjpe, S. Bhabal, and K. Kanere, "Indian sign language converter system using an Android app," in Proc. IEEE Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA), 2017, pp. 436–440. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8212852

[15]. Singh and A. K. Bansal, "An integrated model for automated identification and learning of conversational gestures in human-robot interaction," in Proc. Int. Conf. Comput. Vis. Robot., 2023, pp. 25–36. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-44127-1_3

[16]. Researchgate.net. Retrieved May 7, 2025, from https://www.researchgate.net/profile/Rupesh-Jaiswal/publication/371831685/figure/fig1/AS:11431281170202362@1687639425044/Hand-gesture-recognition-process-of-the-proposed-model.ppm