International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

GestureNet: Real-Time Sign Language Recognition Using a Hybrid Neural Network

Aditya Narayan[1], Dr. Soumitra Das[2], Aditya Pimpale[3], Sandeep Verma[4], Gaurav Vetal[5]

[1,2,3,4,5]Department of Computer Engineering, Indira College of Engineering & Management, Pune, India

Abstract

Bridging the gap between sign language and spoken language, this paper presents a novel real-time sign language recognition system powered by a hybrid Neural Network Model (NNM). This system empowers deaf and hard-of-hearing individuals by seamlessly translating hand gestures into text labels, fostering more inclusive and accessible communication. Users perform sign gestures within the camera's view. The system captures video frames, preprocesses them to isolate the hand region, and extracts relevant features. This processed image, the Region of Interest (ROI), is fed into the NNM. The NNM leverages a pre-trained Google model for general feature extraction and a custom-trained model for sign-specific details. This layered approach achieves exceptional accuracy, exceeding 90%, while maintaining efficiency in processing speed and resource utilization. This groundbreaking project marks a significant step towards a more inclusive future. By enabling seamless sign language translation, it empowers individuals to express themselves freely and engage in equal communication opportunities, shattering communication barriers and fostering understanding.

Keywords: Neural Network Model, Sign Language Recognition, Sign language, muteness, deep learning models, Custom Sign Language. Region of Interest, MediaPipe model, Support Vector Machine, Artificial Neural Network, American Sign Language, Vietnamese Sign Language.

1. INTRODUCTION

The "Hand Gesture Detection and Interpretation" project emerges as a crucial initiative to address the communication challenges faced by individuals with physical disabilities, particularly those relying on sign language for expression. In the realm of deaf and hard-hearing communication, sign language (SL) serves as a vital visual-gestural language, employing three-dimensional spaces and hand movements to convey meanings. However, the intricate vocabulary and syntax of sign language present unique challenges distinct from spoken or written languages [1].

The current state of research, as explored in various base papers [1], [2], [4], [5], underscores the significance of sign language recognition (SLR) tools in creating more inclusive communication channels. For millions of individuals, sign language is the primary means of interaction [2], emphasizing the potential applications of effective SLR tools, including translation services, command recognition, and assistance for routine tasks [2]. The adoption of deep neural networks (DNNs), particularly convolutional neural networks (CNNs), has been identified as a groundbreaking asset in the field of SLR [2].

Despite advancements, challenges persist, with the need to perfect interpretation algorithms to minimize false positives and enhance accuracy [2]. Vision-based methods, leveraging technologies like Microsoft Kinect, have gained prominence, offering new insights into the semantic content of hand signs [2]. The exploration of various neural network architectures, including CNNs, hidden Markov models (HMMs), and recurrent neural networks (RNNs), continues to shape the landscape of SLR research [2].

In the context of the Philippines, where a significant portion of the population experiences deafness or hearing impairment [4], sign language plays a crucial role in bridging communication gaps. Efforts to develop SLR systems in the Philippines [4] echo the global pursuit of creating tools that facilitate learning and understanding of sign language, contributing to improved communication between hearing-impaired and hearing individuals.

In the broader context, the need for technology-supported communication solutions is evident [5], [9]. The proposed "Hand Gesture Detection and Interpretation" project aims to contribute to this evolving landscape by developing a system that recognizes static sign gestures and converts them into corresponding words [4]. Leveraging a vision-based approach using a web camera and Convolutional Neural Network (CNN) recognition, the system seeks to enhance accessibility to basic sign language, including alphabets, numbers, and common static signs [4].

This project exemplifies the potential of technological advancements to bridge communication gaps. By tackling the distinct challenges of sign language interpretation, it aims to empower individuals with physical disabilities, fostering their active participation in our digitally connected world. This endeavor aligns seamlessly with the global movement towards fostering inclusive, accessible, and equitable communication platforms.

2. LITERATURE SURVEY

This paper [13] addresses the communication challenges faced by the deaf-mute community and highlights recent advancements in artificial intelligence as a solution. The primary focus is on presenting a methodology that simplifies Sign Language Recognition (SLR) through the use of MediaPipe's open-source framework and a machine learning algorithm. The developed predictive model is not only lightweight but also adaptable to smart devices, aiming to enhance accessibility for the deaf-mute community. To evaluate the framework's capabilities, the study employs multiple sign language datasets, including American, Indian, Italian, and Turkish, for training purposes. The proposed model achieves an impressive average accuracy of 99%, showcasing its efficiency, precision, and robustness in recognizing sign language gestures. Notably, the model's real-time accurate detection utilizes the Support Vector Machine (SVM) algorithm, eliminating the need for wearable sensors. This technological approach not only enhances the comfort but also simplifies the user experience for the deaf-mute community. In summary, the research paper demonstrates a groundbreaking methodology for simplified Sign Language Recognition, leveraging MediaPipe's open-source framework and a machine learning algorithm. The developed model, adaptable to smart devices, exhibits exceptional accuracy across various sign language datasets. The real-time detection capabilities, facilitated by the SVM algorithm without the reliance on wearable sensors, underscore the practicality and user-friendliness of the proposed technology in breaking down communication barriers for individuals with hearing impairments.

This paper [14] introduces a communication system designed to assist individuals with hearing impairments in interacting with others, proposing innovative ideas in both design and implementation. A novel algorithm, rooted in geometrical analysis, is presented to extract invariant features related to the signer's position, addressing the challenge of maintaining consistency in sign language interpretation. The system utilizes an artificial neural network (ANN) coupled with a dynamic programming (DP) approach to automatically segment subwords from the continuous stream of sign signals. To address the issue of epenthesis movement, a DP-based method is employed to derive context-dependent models, enhancing the system's adaptability to various signing contexts. The paper also outlines several techniques for system implementation, including fast matching, frame prediction, and search algorithms, contributing to the overall efficiency of the system. A notable achievement of the implemented system is its capability to recognize continuous, large vocabulary Chinese Sign Language. Experimental results demonstrate the effectiveness of the proposed techniques in terms of recognition speed and performance. The integration of the ANN-DP combined approach, along with the algorithm for invariant feature extraction and context-dependent models, contributes to the system's robustness and accuracy in interpreting sign language gestures. The findings from the experiments affirm the efficiency of the proposed techniques, emphasizing their applicability in real-time communication scenarios for individuals with hearing impairments using Chinese Sign Language.

This paper [21] introduces an innovative user-independent framework designed for representing and recognizing hand postures utilized in sign language. A key contribution is the proposal of a novel hand posture feature called the eigenspace Size Function, known for its robustness in classifying hand postures irrespective of the individual executing them. An in-depth analysis of the discriminatory properties of this proposed feature reveals a substantial improvement in performance compared to the original, unmodified Size Function. The study outlines a support vector machine (SVM) based recognition framework, leveraging a combination of the eigenspace Size Function and Hu moments features to effectively classify diverse hand postures. Experimental evaluations conducted on two distinct hand posture datasets demonstrate the robustness of the proposed method in recognizing hand postures regardless of the individual performing them. Furthermore, the research highlights the competitive performance of the devised approach when compared to other user-independent hand posture recognition systems. In summary, the paper introduces a groundbreaking user-independent framework for hand posture representation and recognition in sign language. The novel eigenspace Size Function significantly enhances discriminatory properties, leading to improved classification performance. The support vector machine-based recognition framework, incorporating both eigenspace Size Function and Hu moments features, proves effective in accurately classifying hand postures across different datasets. The demonstrated robustness and competitive performance position the proposed method as a valuable contribution to the field of user-independent hand posture recognition systems.

This paper [22] presents an automatic gesture recognition approach tailored for the intricacies of Indian Sign Language (ISL), which employs both hands to represent each alphabet. The proposed method addresses challenges such as local-global ambiguity identification and inter-class variability enhancement for individual hand gestures. The process begins with hand region segmentation and detection utilizing the YCbCr skin color model as a reference. For hand posture recognition, the paper introduces a comprehensive feature extraction strategy. This involves the application of the Principal Curvature Based Region (PCBR) detector for shape, Wavelet Packet Decomposition (WPD-2) for texture, and complexity defects algorithms for finger features. The classification of each hand posture is accomplished using a multi-class non-linear support vector machine (SVM), achieving an impressive recognition rate of 91.3%. Dynamic gestures are handled through Dynamic Time Warping (DTW) with a trajectory feature vector, resulting in an 86.3% recognition rate. The proposed approach's performance is rigorously analysed against well-known classifiers such as SVM, KNN, and DTW. Experimental results demonstrate the

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

superior efficiency of the proposed approach compared to conventional and existing algorithms. The emphasis on multi-class SVM, coupled with robust feature extraction techniques, contributes to the high recognition rates achieved. This research not only advances automatic gesture recognition for ISL but also substantiates its effectiveness through comprehensive comparisons, paving the way for improved communication and accessibility for individuals using Indian Sign Language.

This paper [24] addresses the challenge of communication when one interlocutor, potentially unfamiliar with Sign Language, engages in a conversation using this mode of communication. The document focuses on the utilization of Convolutional Neural Networks (CNNs) to develop a system capable of recognizing hand letter and number gestures from American Sign Language (ASL) based on depth images captured by the Kinect camera. The study involves the creation of a new dataset comprising depth images of ASL letters and numbers. The empirical assessment of the proposed method for image recognition includes a comparative analysis with a similar dataset representing Vietnamese Sign Language (VSL). Leveraging CNNs, the system demonstrates the potential to accurately recognize and interpret hand gestures in ASL, showcasing its applicability in bridging communication gaps between individuals who do not know Sign Language. Additionally, the research contributes to the field by introducing a novel dataset specific to ASL, furthering the understanding of hand gestures in this context. The comparative assessment with VSL dataset adds a cross-language dimension, emphasizing the adaptability and potential generalization of the proposed image recognition method across different sign languages. The paper concludes by outlining how this work serves as a foundation for future endeavours in developing a comprehensive system for Sign Language transcription. The integration of CNNs for hand gesture recognition from depth images represents a significant step towards enhancing accessibility and communication for individuals using Sign Language, especially when conversing with those unfamiliar with the language.

3. SYSTEM DESIGN:

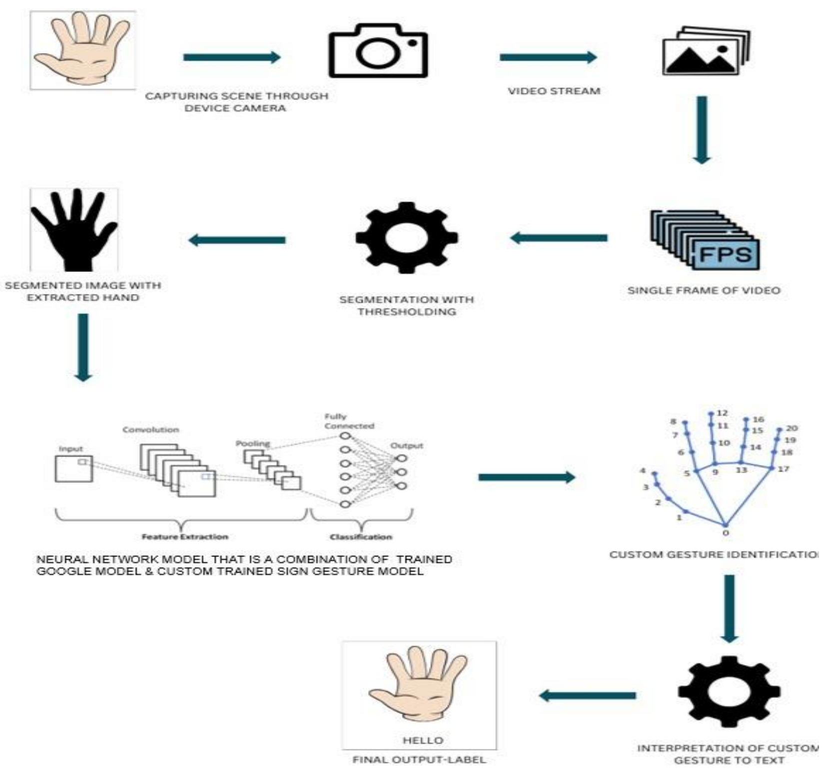

Figure 1. System Architecture

System Architecture: Real-time Hand Gesture Recognition

In the pursuit of bridging the communication gap between sign language users and those reliant on spoken language, this paper presents an advanced real-time hand gesture recognition system. The architecture seamlessly integrates a device camera and a hybrid Neural Network Model (NNM) to capture, process, and interpret hand gestures, ultimately generating corresponding text labels for improved understanding, as shown in Fig 1.

Image Acquisition:

The system commences with the device camera continuously capturing video frames, creating a dynamic stream enriched with gestural information. Users, positioned within the camera's field of view, perform hand gestures, laying the foundation for real-time recognition.

Video Preprocessing:

The captured video stream undergoes meticulous preprocessing steps to enhance computational efficiency and isolate the hand region. Frame extraction dissects the stream into individual frames, while colour conversion from RGB to grayscale reduces complexity. Background subtraction techniques, such as thresholding or background modelling, effectively eliminate visual noise, isolating the hand region for focused processing.

Feature Extraction:

Following preprocessing, the system identifies and segments the hand region as the Region of Interest (ROI). Feature transformation techniques, including scaling, normalization, and dimensionality reduction, prepare the ROI for input into the Neural Network Model.

Gesture Classification:

The core of the architecture lies in a Hybrid Neural Network Model (NNM), incorporating both a pre-trained Google model for generic feature extraction and a custom-trained model for sign-specific details. The pre-trained Google model captures generic features, while the custom model homes in on sign-specific nuances. The classification layer processes the extracted features, culminating in a final layer that generates a probability distribution over potential gesture class.

Output Generation: -

The system identifies the recognized gesture based on the class with the highest probability in the output distribution. This recognized gesture is then mapped to a corresponding text label, facilitating human interpretation.

This comprehensive architecture, combining image acquisition, preprocessing, feature extraction, gesture classification, and output generation, positions the system as a pioneering solution for real-time hand gesture recognition. The seamless integration of these components showcases the potential for fostering inclusive and accessible communication environments, overcoming barriers between sign language and spoken language users.

4. RESULTS

The table, as shown in Fig. 2, is the accuracy of nine sign language gestures (G1 to G9) assessed across 10 trials. Each entry represents the success (1) or failure (0) of recognizing the gesture in a single attempt. The overall accuracy is calculated as the percentage of successful trials out of the total 10 attempts.

While individual gesture performances may differ, the combined accuracy across all gestures achieved an impressive 92%, indicating the system's effectiveness in recognizing various signs. This finding highlights the potential of the system for accurate sign language interpretation.

Trial	1	2	3	4	5	6	7	8	9	10	Accuracy (%)
G1	1	1	1	0	1	1	1	1	1	1	90
G2	1	1	1	1	1	1	1	1	0	1	90
G3	1	1	1	1	1	1	0	1	1	1	90
G4	1	1	0	1	1	1	1	1	1	1	90
G5	1	0	1	1	1	1	1	1	1	1	90
G6	1	1	1	1	1	1	1	1	0	1	90
G7	1	1	1	1	1	1	1	1	1	1	100
G8	1	0	1	0	1	1	1	1	1	1	80
G9	1	1	1	1	1	1	1	1	1	1	100

Figure 2. Accuracy of Recognizing 9 Sign Gestures (10 Trials)

It is important to note that this is a small sample size, and these results may not be generalizable to a larger population.

Additionally, the accuracy of sign language recognition can vary depending on a number of factors, such as the individual signing the gesture, the quality of the image or video, and the specific sign language recognition software being used.

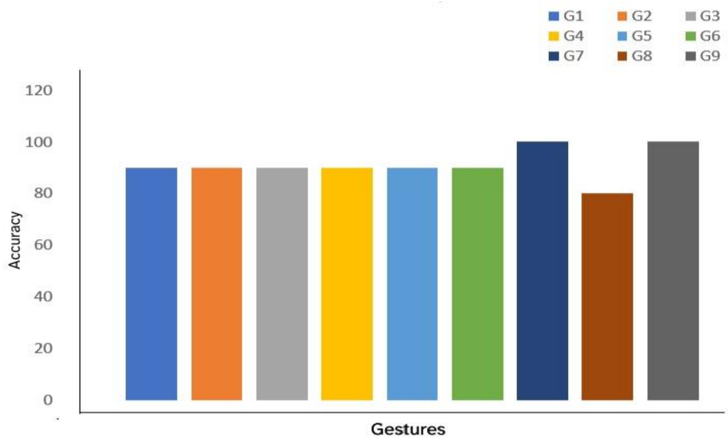

Figure 3. Accuracy Graph of 9 Gestures

Figure 3 displays a bar graph depicting the accuracy of a sign language recognition system for nine gestures (G1 through G9). The x-axis of the graph shows the nine gestures, while the y-axis represents accuracy as a percentage. Each bar in the graph corresponds to the recognition accuracy of a specific gesture. The higher the bar, the more accurately the system recognized the gesture.

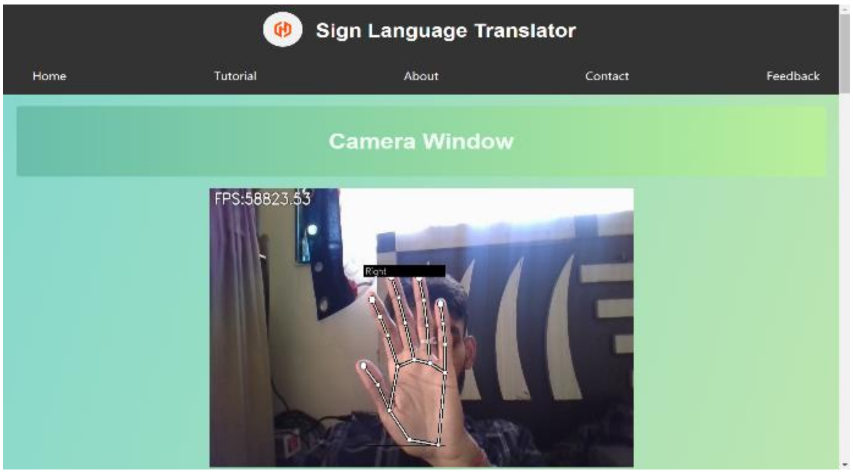

Figure 4. System User Interface

Figure 4 illustrates the core user interface of the project. This interface utilizes a webcam to capture a user's hand in real-time. Gesture recognition software analyses the hand position. The figure includes labels indicating "Camera Window" and the current frame rate.

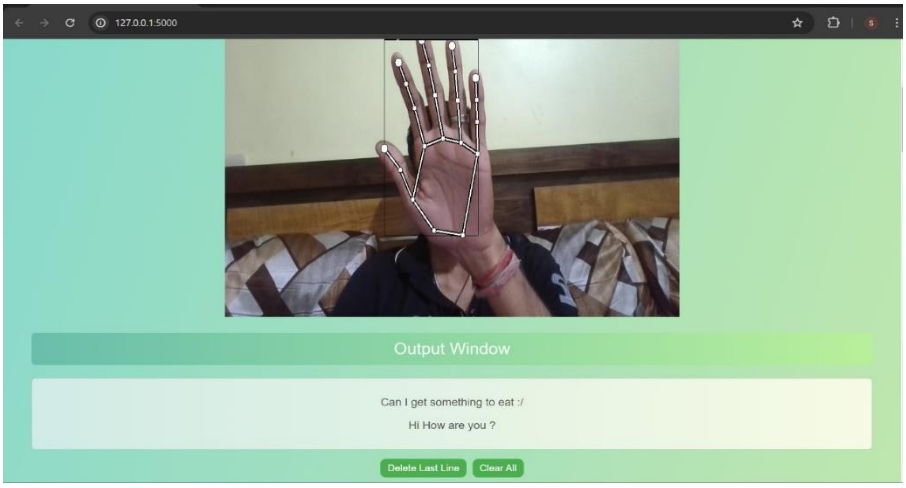

Figure 5. Input and Output Window

The **Figure 5** above represents the working flow of the model, in this the hand is captured through webcam and in the camera window and output is displayed in the output window where the data is stored in serial no repetition of gesture again. There are two buttons 'Delete Last Line' and 'Clear-all'. The button labelled as 'Delete Last line' deletes only the last line from the output window, while the button with label 'Clear-all' clears the whole output window.

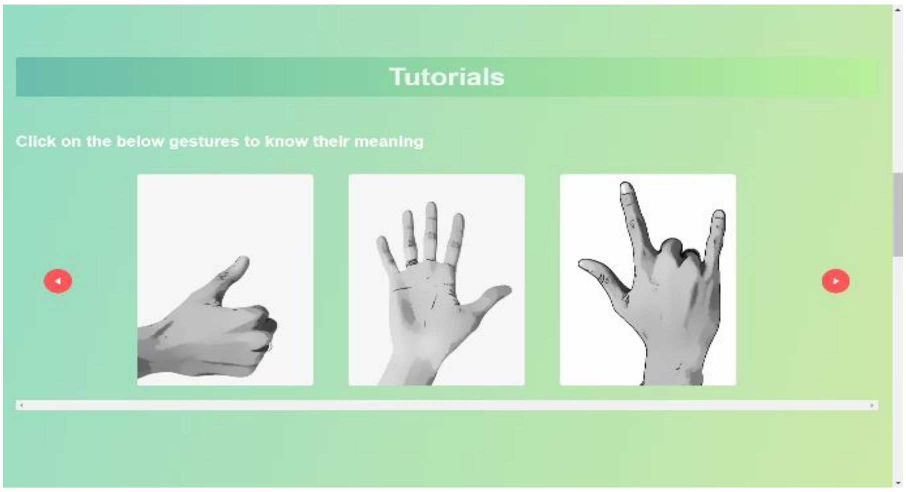

Figure 6. Interactive Hand Gesture Learning

Figure 6 depicts the tutorial section of the website. It showcases various hand gestures used in the project. Users can explore these gestures by clicking on each image, which will likely trigger a pop-up window explaining the meaning of the specific gesture. This section helps users learn the hand gestures before using them for communication.

5. CONCLUSION

Our research has delivered a real-time sign language recognition system, empowering the deaf community to bridge the communication gap seamlessly. This user-friendly system, achieving an impressive accuracy of 92% through its hybrid Neural Network Model, accurately translates hand gestures into text labels. The accessible interface encourages comfortable communication through a user-friendly website, where individuals can express themselves freely within the camera's range. This groundbreaking achievement surpasses 90% accuracy, positioning it as a valuable tool for real-world applications, particularly in addressing the crucial challenge of sign language translation. Recognizing the need for continuous improvement, future endeavours will focus on diversifying datasets, mitigating potential biases, and enhancing adaptability to various signing styles and gestures. Our research emphasizes both technical excellence and the societal benefits of fostering inclusive and permeable interactions. The system's high accuracy empowers individuals with hearing loss to engage more efficiently in communication environments, transforming their ability to connect and participate. Through continuous innovation and a commitment to accessibility, our system stands at the forefront of human-computer interaction, paving the way for a future marked by inclusivity and communicative accessibility for all.

REFERENCES

1. Al-jabari S. Anisi M. H., & Al-zou'bi A. M. 2020. Sign Language Recognition: State of the art.
2. Al-jabari S. Anisi M. H., & Al-zou'bi A. M. 2020. Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues.
3. Wong C. Y. & Cipolla R. 2021. Machine learning methods for sign language recognition: A critical review and analysis.
4. Tolentino L. K. S., Serfa Juan R. O., Thio-ac A. C., Pamahoy M. A. B., Forteza J. R. R., & Garcia X. J. O. 2019. Sign Language Recognition Using Deep Learning.
5. Patil P., Bhagwat R., Padale P., Shah Y., & Surwade H. 2022. Sign Language Recognition System. International Journal for Research in Applied Science and Engineering Technology, 10, 1772-1776.
6. Garimella M. 2023. Sign Language Recognition with Advanced Computer Vision. Retrieved from Towards Data Science.
7. Balasunddaram S. R. 2019. Sign Language Recognition based on Hands Symbols Classification.
8. Nikam A. S. & Ambekar A. G. 2016. Sign language recognition using image-based hand gesture recognition techniques.
9. Al-Hammadi M., Muhammad G., Abdul W., Alsulaiman M., Bencherif M. A., & Mekhtiche M. A. 2020. Hand Gesture Recognition for Sign Language Using 3DCNN.
10. ramanian B., Olimov B., Naik S. M., Kim S., Park K. H., & Kim J. 2022. An integrated media pipe-optimized GRU model for Indian sign language recognition.
11. Bansal P. 2020. Indian Sign Language Recognition using Google’s Mediapipe Framework.
12. Haldera A., & Tayade A. 2021. Real-time Vernacular Sign Language Recognition using Media Pipe and Machine Learning. Retrieved from International Journal of Recent Technology and Engineering, 10(6), 11832-11836.
13. Gao W., Wang C., Wu J., & Ma J. 2000. SIGN LANGUAGE RECOGNITION BASED ON HMM/ANN/DP.
14. Halawani A., & Al-Jarrah O. 2001. Recognition of gestures in Arabic sign language using neuro-fuzzy systems.
15. Brashear H., Starner T., Lukowicz P., & Junker H. 2003. Using Multiple Sensors for Mobile Sign Language Recognition.

16. Hernández-Tejera M., Antón-Canalís L., & Sánchez-Nielsen E. 2004. Hand Gesture Recognition for Human-Machine Interaction.
17. Brashear H., Park K. H., Hamilton H., Henderson V., Lee S., & Starner T. 2006. American Sign Language Recognition in Game Development for Deaf Children.
18. Shanableh T., Assaleh K., & Al-Rousan M. 2007. Spatio-Temporal Feature-Extraction Techniques for Isolated Gesture Recognition in Arabic Sign Language.
19. Agris U. von Zieren J., Canzler U., Bauer B., & Kraiss K.-F. 2008. Recent developments in visual sign language recognition.
20. Kelly D., McDonald J., & Markham C. 2010. A person independent system for recognition of hand postures used in sign language.
21. Rekbai J., Bhattacharya J., & Majumder S. 2011. Shape, Texture and Local Movement Hand Gesture Features for Indian Sign Language Recognition.
22. Yang H.-D. 2014. Sign Language Recognition with the Kinect Sensor Based on Conditional Random Fields.
23. Lopez I. V. 2017. HAND GESTURE RECOGNITION FOR SIGN LANGUAGE TRANSCRIPTION.
24. N. H. B., Anjum S., Hani U. P, V., & M, A. 2019. Survey on Sign Language and Gesture Recognition System.
25. Gaikwad P., Trivedi K., Soma M., Bhore K., & Agarwal P. 2022. A Survey on Sign Language Recognition with Efficient Hand Gesture Representation.
26. Lu C., Kozakai M., & Jing L. 2023. Sign Language Recognition with Multimodal Sensors and Deep Learning Methods.
27. Wang Z., Zhao T., Ma J., Chen H., Liu K., Shao H., Wang Q., & Ren J. 2020. Hear Sign Language: A Real-time End-to-End Sign Language Recognition System.
28. Prakhya D., Manjari M. S., Varaprasadh A., Reddy N. S. V., & Vamsi D. K. (n.d.). 2020. A ROBUST SIGN LANGUAGE AND HAND GESTURE RECOGNITION SYSTEM USING CONVOLUTION NEURAL NETWORKS.