International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 13 Issue 4, April, 2024, Impact Factor: 8.285

"Sign Wave: A Compassionate Companion for the Hearing Impaired"

Yadav Niraj[1], Shewale Jayesh[2], Waghmare Pranav[3], Shraddha Subhedar[4]

[1,2,3,4] Department of Computer Science Engineering, Artificial Intelligence and Machine Learning, Saraswati College of Engineering Kharghar, India

ABSTRACT

In recent years, advancements in computer vision and machine learning have paved the way for the hearing-impaired community, offering solutions that enhance communication and accessibility. This paper introduces an innovative approach to hand sign gesture recognition, utilizing Google MediaPipe for precise landmark detection, thereby enhancing model training efficiency and accuracy. Departing from complex Convolutional Neural Networks (CNNs), our method offers a simpler implementation process while maintaining high performance. By leveraging Google MediaPipe, our system identifies landmarks within hand sign images, improving recognition accuracy. What distinguishes our project is its dual focus on gesture recognition and bidirectional communication. We not only recognize hand gestures but also enable two-way communication by translating text inputs into corresponding hand signs, fostering enhanced accessibility. Through comparative analysis, we demonstrate the effectiveness of our method in hand sign gesture recognition, aiming to bridge communication gaps and empower individuals within the deaf community. Our approach promises to revolutionize gesture recognition while promoting inclusivity and accessibility.

Index Terms - Bidirectional Communication, Convolutional Neural Networks (CNNs), Google MediaPipe, Hand Gesture Recognition, Indian Sign Language (ISL)

INTRODUCTION

The ability to communicate effectively transcends barriers and unites humanity. For the millions navigating the world with hearing impairments, sign languages serve as an indispensable bridge, fostering inclusion and connection. However, formidable obstacles persist, including a dearth of qualified interpreters, the intricacy of mastering sign languages, and limited access to communication aids. Our research aims to innovate hand gesture recognition for Indian Sign Language (ISL), pivotal in overcoming these challenges.

While traditional methods rely on Convolutional Neural Networks (CNNs), their implementation hurdles and limited generalization across sign languages pose significant limitations. Our alternative approach leverages Google MediaPipe for precise hand landmark detection, streamlining training while enhancing recognition accuracy.

Beyond recognition, we endeavor to enable bidirectional communication by translating text inputs into sign language representations, thereby expanding accessibility horizons. To substantiate our work, we conduct a comprehensive review of existing sign language recognition systems, emphasizing those tailored for ISL. Through comparative analysis, we provide empirical evidence highlighting the efficacy of our landmark-based approach in fostering accessible sign language communication. This research contributes to advancing inclusive communication for individuals with hearing impairments.

LITERATURE SURVEY

Introducing their system, the authors of [1] propose a solution to aid Portuguese Sign Language users by converting sign language gestures into 3D avatar animations. However, they overlook the needs of Indian Sign Language users and lack bidirectional communication capabilities. To address this, they suggest customizing motion capture technology for Indian Sign Languages, integrating OpenCV and CNN for gesture detection, and developing a tailored 3D avatar model.

In their groundbreaking work, researchers from Pune, India, in [4], present a method to convert Indian Sign Language into speech using Deep Neural Networks. Leveraging CNNs with Google's Text-to-Speech API, the system achieves an impressive 98% accuracy rate. The authors emphasize the need to adapt technologies specifically for Indian Sign Languages and implement bidirectional communication functionalities.

The authors of [5] introduce a three-module system aimed at facilitating communication between Indian Sign Language (ISL) users and English speakers. By utilizing IBM-Watson and regular expressions, spoken English is converted into ISL sentences, represented by a 3D avatar. They stress the importance of customization to suit Indian Sign Languages and the development of tailored evaluation criteria.

Introducing "Signet," in [7], the authors employ Convolutional Neural Networks to recognize static alphabet signs in Indian Sign Language (ISL). While focused on ISL recognition, the system lacks consideration for broader Indian Sign Languages and does not include text-to-sign conversion for bidirectional communication. They propose expanding the scope to encompass diverse Indian Sign Languages and incorporating text-to-sign conversion functionalities.

Table 1 outlines key research in sign language processing, covering animation capture, sign-to-speech via deep neural networks, 3D avatar development for sign movement, and deep learning-based recognition systems.

Table I.
Sign Language Survey

Sr. No.	**Year**	**Title**	**Author Name**	**Journal Name**	**Research**
1	2023	Capturing and Processing Sign Animations to a Portuguese Sign Language 3D Avatar	Bruno Ribeiro, Duarte Dias, et al	ACM Transactions on Graphics (TOG), IEEE Conference Publication	Deaf, Assistive Technology, PSL, 3D Avatar, Animation, 3D Models, Motion Capture
2	2022	Conversion of Indian Sign Language to Speech by Using Deep Neural Network	Sonali Patil, Shivam Gulave, et al	IEEE Conference Publication	Indian Sign Language, Sign to text conversion, Hand gestures, CNN, DNN.
3	2021	3D Avatar Approach for Continuous Sign	Pradeep kumar, et al	Applied Sciences	Indian Sign Language (ISL), Natural language processing,

Table I.
Sign Language Survey

Sr. No.	Year	Title	Author Name	Journal Name	Research
		Movement Using Speech/Text			Avatar, Sign movement.
4	2019	Signet: A Deep Learning based Indian Sign Language Recognition System	Sruthi C. J and Lijiya A.	IEEE Conference Publication	Convolutional Neural Network, Deep Learning, Gesture recognition

PROPOSED SYSTEM

Our approach main focus on following objectives

1) Develop a hand sign recognition system utilizing Google MediaPipe for improved accuracy and efficiency.
2) Enable bidirectional communication by converting text inputs into hand sign representations to enhance accessibility.

Architecture

Following the architecture of SignWave Hand sign gesture recognition system and SignWave bidirectional communication.

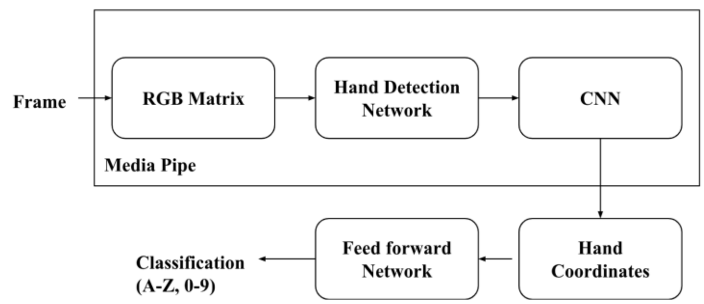

Fig 1. Architecture of SignWave Hand sign gesture recognition system

The proposed approach involves real-time hand gesture recognition utilizing Google MediaPipe and an image classification module, as illustrated in Fig 1.

1) Image Acquisition: OpenCV capture frames from the camera, providing real-time video input.
2) Color Space Conversion: RGB images undergo conversion to ensure compatibility with subsequent processing.
3) Hand Detection: Google MediaPipe's hand detection network locates and localizes hand regions in captured frames.
4) Landmark Detection: MediaPipe extracts hand landmarks, offering spatial information crucial for gesture recognition.

5) Hand Gesture Representation: A Convolutional Neural Network (CNN) processes the extracted hand landmarks, undergoing preprocessing for accurate classification.
6) Classification: A feedforward neural network analyzes hand coordinates from MediaPipe for gesture classification.
7) Visualization: Recognized hand gestures are displayed in real-time, providing users with visual feedback for enhanced interaction.

This streamlined process aims to improve accuracy and efficiency in hand sign recognition, enhancing accessibility for individuals with hearing impairments.

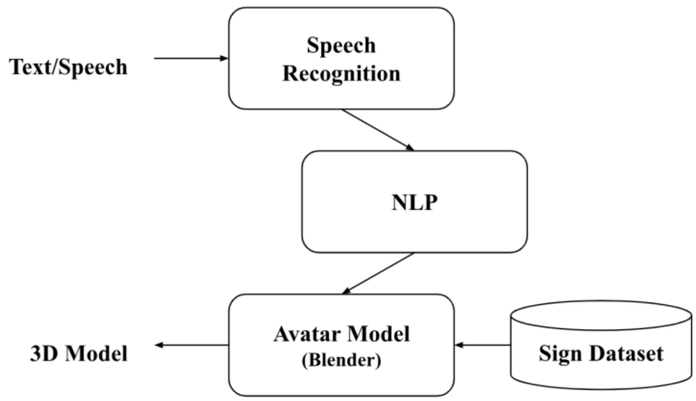

Fig 2. Architecture of SignWave bidirectional communication

Bidirectional communication by translating user-provided text or speech inputs into corresponding SignWave sign language gestures. The process involves several steps, including speech recognition, natural language processing (NLP), and integration with an avatar model capable of rendering SignWave sign language gestures in Fig 2.
1) Input Acquisition: The SignWave system receives input from the user, either in the form of text or speech.
2) Speech Recognition: If the input is speech, a speech recognition system converts it into text format.
3) Natural Language Processing (NLP): The transcribed text undergoes SignWave NLP analysis to extract semantic meaning and linguistic structure.
4) Avatar Model Blender: The processed text is integrated with a SignWave avatar model blender, which combines NLP output with pre-existing sign language gestures.

Gesture Recognition and Visualization: The blended model generates visual representations of SignWave sign language gestures corresponding to the input text, enabling bidirectional communication.

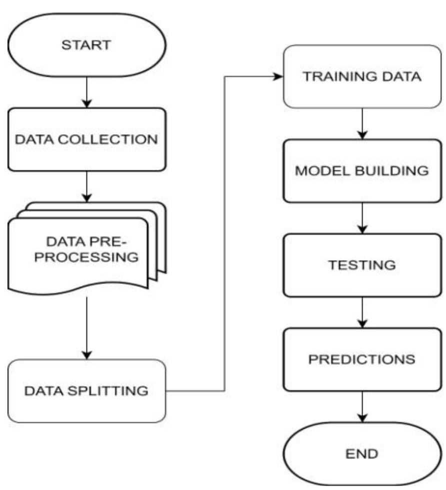

Fig 3 Flow of the SignWave System

In the SignWave project, data collection encompasses surveys, experiments, and public databases, followed by meticulous pre-processing to ensure quality. The data is then partitioned into training and testing sets to train and evaluate the model respectively. Model selection involves choosing a machine learning algorithm trained on the provided data, with subsequent evaluation using testing data. Successful models are then deployed for predictions, as depicted in Fig 3, showcasing SignWave's systematic approach towards robust machine learning solutions.

RESULT AND ANALYSIS

SignWave Project: Achieving Recognition Success and Enhancing Dataset Diversity.

Dataset Creation
Creating an effective dataset poses a significant challenge for neural networks, particularly in domains like sign language recognition. In our research, we encountered a scarcity of comprehensive datasets tailored to real-world conditions. Existing datasets often lacked diversity, containing similar images that didn't accurately reflect the variability found in real-time scenarios, such as camera noise and variations in brightness.

Recognizing this limitation, we embarked on the creation of our own dataset comprising over 3,000 images of hand gestures. This dataset addresses the shortcomings of existing ones by incorporating a diverse range of real-world conditions. Each class, representing letters from A to Z and digits from 0 to 9, comprises more than 1,000 images. By curating such a robust dataset, we aim to enhance the model's ability to generalize and perform effectively in real-time prediction tasks.

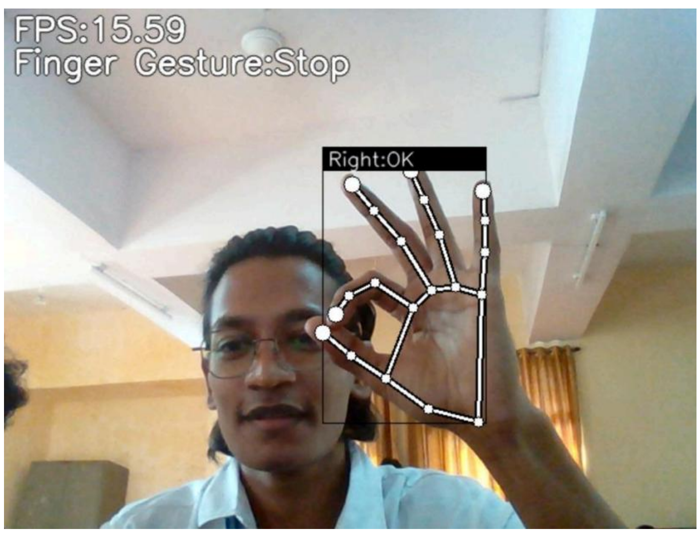

Fig. 4. Glimpse of SigWave System

The Fig. 4 is a glimpse of a proposed system's result. "Finger Gesture: Open". This suggests that the System is successfully recognizing a gesture from the user's hand.

SignWave is analyzed against three parameters. The precision, recall and f1-score is calculated using the formulas:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{1}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{2}$$

$$f1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3}$$

The high values of precision, recall, and F1-score signify a good classification result, indicating the effectiveness of the proposed method. Comparative analysis against existing methods reveals that the proposed method outperforms others in terms of dataset size and accuracy.

TABLE II.
CLASS-WISE PERFORMANCE METRICS

Class	Precision	Recall	F1-score
A	0.36	0.28	0.38
B	0.38	0.38	0.38
C	0.48	0.58	0.58
D	0.43	0.43	0.43
E	0.59	0.59	0.59
F	0.56	0.56	0.56
G	0.46	0.46	0.46
H	0.54	0.64	0.64
I	0.65	0.55	0.55
J	0.55	0.65	0.65
K	0.59	0.49	0.49
L	0.57	0.57	0.57
M	0.52	0.42	0.42
N	0.48	0.52	0.42
O	0.57	0.47	0.47
P	0.46	0.56	0.56
Q	0.57	0.36	0.26
R	0.44	0.54	0.64
S	0.44	0.54	0.64
T	0.49	0.49	0.49
U	0.46	0.46	0.46
V	0.47	0.47	0.47
W	0.46	0.46	0.46
X	0.47	0.57	0.57
Y	0.42	0.47	0.47
Z	0.57	0.52	0.42
1	0.88	0.67	0.67
2	0.81	0.83	0.82
3	0.80	0.87	0.83
4	0.60	0.68	0.64
5	0.89	0.84	0.86
6	0.60	0.65	0.65
7	0.60	0.65	0.65
8	0.40	0.46	0.46
9	0.43	0.47	0.47

Table 2, Class-wise Precision, Recall, and F1-score obtained from model predictions and ground truth labels. Each row represents a class, and corresponding metrics are calculated to evaluate classification performance.

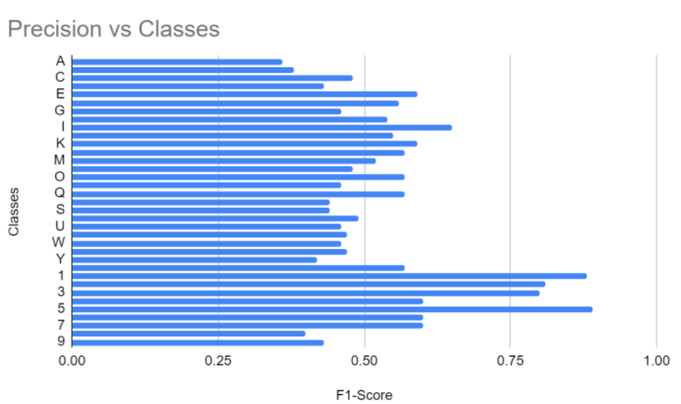

Fig. 5. Precision Vs Classes.

The precision obtained after testing reflects the classifier's accuracy in predicting true positive samples. Fig. 5 depicts the precision results, showcasing the model's precision performance.

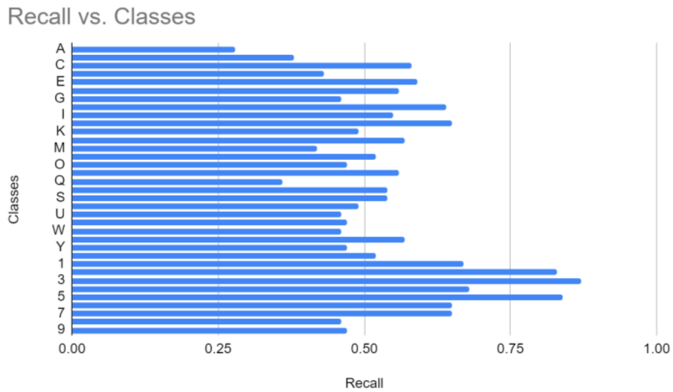

Fig. 6. Recall Vs Classes.

Recall, on the other hand, measures the proportion of actual positive samples that the classifier correctly identifies. Fig. 6 illustrates the recall results, representing the model's recall performance.

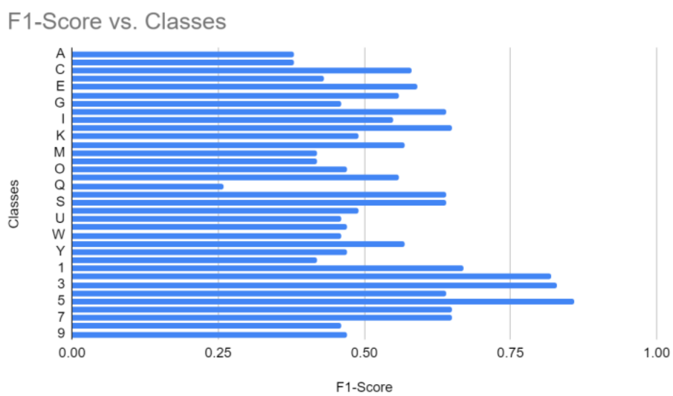

Fig. 7. F1-Score Vs Classes

The F1-score, derived as the harmonic mean of precision and recall, offers a comprehensive assessment of classification performance. Fig. 7 presents the F1-score results, reflecting the overall effectiveness of the model in classifying the dataset.

SCOPE AND FUTURE WORK

This paper focuses on developing and evaluating a hand sign gesture recognition system that utilizes Google MediaPipe for landmark detection and Convolutional Neural Networks (CNNs) for classification. The scope covers recognizing hand gestures from the Indian Sign Language (ISL) alphabet (0-9, A-Z). While the MediaPipe component has been implemented, generating a 3D avatar for visual hand sign representation remains incomplete, limiting bidirectional communication capability. However, a natural language to sign language translation module facilitates text-to-sign conversion.

The lack of a 3D avatar hinders effective visual communication for individuals with hearing impairments, impeding comprehensive accessibility. Addressing this limitation and achieving bidirectional functionality requires further development efforts, potentially involving collaboration with 3D modeling experts. The system's performance is evaluated through accuracy, precision, recall, and F1 score metrics, with comparative analysis against existing approaches demonstrating its efficacy and efficiency.

Future work will focus on overcoming the 3D avatar generation limitation, incorporating visual representation for bidirectional communication, and continuously improving the gesture recognition system's accuracy and usability.

CONCLUSION

In conclusion, this study presents a promising approach to hand sign gesture recognition, employing Google MediaPipe for landmark detection and Convolutional Neural Networks (CNNs) for classification. The system effectively recognizes hand gestures representing the Indian Sign Language (ISL) alphabet (0-9, A-Z).

While currently lacking a 3D avatar for visual representation, our methodology lays the foundation for future advancements in bidirectional communication solutions. By leveraging advanced technologies and deep learning techniques, this system has the potential to significantly enhance communication accessibility for individuals with hearing impairments.

In summary, this work contributes positively to ongoing efforts to improve communication accessibility for individuals with hearing impairments. Through the utilization of technology-driven solutions, we aim to foster inclusivity and empower marginalized communities, highlighting the transformative possibilities of our approach.

REFERENCES

[1] Bruno Ribeiro et al, Capturing and Processing Sign Animations to a Portuguese Sign Language 3D Avatar, 2023 (https://ieeexplore.ieee.org/document/10180233)
[2] J.S.K. Gowda and M.N. Shanmukha Swamy, "Dynamic Hand Gesture Recognition: A Literature Review," AIP Conference Proceedings 2494, 030005, 2023.
[3] Sonali Patil et al, Conversion of Indian Sign Language to Speech by Using Deep Neural Network, 2022 (https://ieeexplore.ieee.org/document/10011043)
[4] Lazzat Zholshiyeva et al., "Hand Gesture Recognition Methods and Applications: A Literature Survey," 7th International Conference on Engineering & MIS 2021(ICEMIS'21), pp. 1-8, 2021.
[5] Pradeep kumar et al, 3D Avatar Approach for Continuous Sign Movement Using Speech/Text, 2021 (https://ieeexplore.ieee.org/document/8698006)
[6] Z. Wang et al., "A Survey on Hand Gesture Recognition for Sign Language Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence 42,1141-1155, 2020.
[7] Sruthi C. J and Lijiya A, Signet: A Deep Learning based Indian Sign Language Recognition System, 2019 (https://ieeexplore.ieee.org/document/8698006)
[8] Ayesha Hameed et al., "Hand Gesture Recognition: A Literature Review," International Journal of Computer Science and Engineering 3, 7, 2013.
[9] S.P. Selvakumar and R. Rajesh, "Hand Gesture Techniques for Sign Language Recognition: A Literature Survey," International Journal of Technical Research & Science (IJTRS) 4, 10, 2014.
[10] V. Keskin et al., "Real-Time Hand Gesture Recognition Using Deep Learning," IEEE Signal Processing Magazine 35, 143-154, 2018.
[11] Julie Bakken Jepsen et al., "Hand Gesture Recognition Methods and Applications: A Literature Survey," in Sign Languages of the World, Walter de Gruyter, Inc., Berlin, pp.283-325, 2015.
[12] D.L. Chen and C.S. Chen, "A Survey on Hand Gesture Recognition," Pattern Recognition 37, 1248-1259, 2004.
[13] I. Essa and T. Darrell, "Hand Gesture Recognition: A Review," Proceedings of the IEEE 86, 1019-1032, 1998.
[14] https://www.researchgate.net/publication/281434972_3D_Animation_framework_for_sign_language