Sign to Speech: A Machine Learning Approach for Deaf and Mute Communication

[1] Dr. Zubin Bhaidasna, [2]Dr. Hetal Bhaidasna

[1]Assistant Professor, [2]Assistant Professor
[1]Department of Computer Engineering,
[1]GCET, CVM University, V. V. Nagar, Gujarat, India

***Abstract*— This research demonstrates a novel attempt to help people who are both deaf and mute by creating a communication assist system that translates hand signs into words. The system uses a camera to capture hand movements and the trained recognition model identifies them. After recognition, text translation followed by speech synthesis through a voice module is performed. To train and evaluate the system, a custom dataset capturing common gestures was created. The sign-to-speech solution is tailored to operate on constrained, cost-effective hardware such as smartphones and tablets. Furthermore, this review discusses the commonly used datasets in sign-to-speech research and their limitations in terms of size, diversity, and standardization. It also suggests a general flow of implementation starting from data collection, preprocessing, feature extraction, model training, and conversion to speech. The paper highlights key challenges such as gesture variability, occlusion, and real-time processing.**

***Index Terms*— Sign Language, Machine Learning, CNN**

I. INTRODUCTION

Communication is an essential human want, yet tens of millions of deaf and mute individuals face every day barriers in expressing themselves, in particular in environments in which sign language isn't widely understood. Sign language, even though wealthy and expressive, stays strange to a great deal of the overall population, developing an opening between hearing-impaired individuals and the rest of society. Traditional answers like interpreters or written communication regularly fall brief in spontaneity and inclusivity, especially in real-time interactions. This disconnects limits get admission to essential offerings, schooling, and social inclusion for the listening to and speech impaired.

To overcome this mission, there is a growing need for structures, which can translate signal language into spoken words, supporting to bridge the distance in conversation. “Sign to Speech” is a step toward allowing smoother interactions by means of converting hand gestures into audible language. Using visual inputs like stay hand actions, the system techniques gestures via structured steps such as video capture, frame evaluation, gesture popularity, and output generation. By focusing on local sign dialects and body motion styles, the system enhances usability and attractiveness among various users. This practical approach empowers individuals to engage more confidently in day-to-day conversations, fostering inclusion and dignity in communication.

II. BACKGROUND STUDY

Effective communique is a center element of human existence, allowing individuals to specific mind, emotions, and needs. For humans with listening to or speech impairments, however, communication can emerge as an extensive barrier, mainly while interacting with the ones unexpected with signal language. Sign language serves as a crucial medium that empowers these people to engage in meaningful communicate. Despite its importance, a verbal exchange divide stays among sign language customers and the general public, primarily due to the lack of giant information of sign language. This hole poses demanding situations in regular situations which includes searching for assistance, accessing services, or participating in social and academic environments, ultimately affecting the overall first-rate of existence of in another way-abled people.

The goal of this studies is to bridge the communication barrier confronted by using individuals with hearing and speech impairments in a greener and cost-powerful manner. A review of existing literature [1] reveals that current methodologies face significant limitations in two critical aspects: recognition accuracy and suitability for real-time implementation. Although a variety of machine learning and deep learning techniques—such as Convolutional Neural Networks (CNN) [2–6], Support Vector Machines (SVM), and K-Nearest Neighbors (KNN)—have been explored in past studies, they often fall short in delivering consistently high performance and responsiveness necessary for practical applications.

Moreover, sure tactics rely on specialized external hardware, including sensor-integrated gloves, which not handiest growth the fee of deployment but additionally lessen the benefit of use for the end user. These demanding situations restriction the accessibility and scalability of such structures, specifically in resource-restrained settings. To overcome those troubles, the present work proposes a solution that extensively complements both the accuracy and actual-time processing capabilities of Indian Sign Language (ISL) reputation systems. The proposed technique is designed to paintings with standard digital camera input, putting off the need for external gadgets, thereby offering an low-cost and user-pleasant opportunity for effective verbal exchange guide [7].

A green actual-time imaginative and prescient-based totally system became advanced for recognizing American Sign Language (ASL) alphabets. One most important assignment became the dearth of a appropriate dataset, leading the researchers to create their very own. The version required square snap shots, just like how CNN in Keras methods statistics. Another challenge was selecting the right image filter out for feature extraction. After testing numerous alternatives, Gaussian blur changed into selected for most excellent results [9]. This look at employs the CamShift set of rules to isolate the human hand from complicated backgrounds and stumble on actual-time gestures. A convolutional neural network then recognizes ten commonly used digit gestures, trained on a dataset of 1,600 hand images (400 per category, totaling 4,000 gestures). The system achieves an impressive 98.3% accuracy in real-time recognition [10]. : In this, a method based on artificial neural networks was suggested for recognizing Indian sign language. It proposed a technique to find a 32-set of combinations—10 for each up and down movement of the fingers—in order to get matching Tamil letters. The method required converting decimal numbers from the up/down position of the fingers into categories that could recognize Tamil alphabets. A collection of static data was recorded as 640 x 480-pixel pictures. Images were converted from RGB to greyscale using palm image extraction [13].

Table 1 Comparison of various sign language Methods

Sign Language	***Dataset***	***Approach Used***	***Methodology***	***Working On***
American Sign Language (ASL)[2]	ASL datasets	Desktop application for real-time sign	CNN	Work on Real-time text conversion
Chinese Sign Language[5]	Custom dataset	Wearable sign language recognition system	CNN, stretchable strain sensors	Combines multiple sliding windows for sentence recognition
American Sign Language (ASL) & British Sign Language (BSL)[6]	Custom dataset	Recognition using computer vision algorithms and neural networks	CNN	Real-time feedback integration, flexible for various settings
Assamese Sign Language[7]	Custom dataset	Recognition system for Assamese Sign Language	MediaPipe, feed-forward neural network	Emphasizes local sign language dialects and utilizes Media Pipe for accurate real-time hand and body landmark detection to interpret gestures

Sign Language	*Dataset*	*Approach Used*	*Methodology*	*Working On*
				effectively.
American Sign Language (ASL)[9]	Custom dataset	Recognize using the Filter CNN	CNN	Works on static dataset
Indian Sign Language (ISL) [16]	Custom dataset	Recognition of ISL gestures	3D CNN	works on both static and dynamic gestures, diverse dataset conditions

III. Methodology of the Proposed Work

The proposed system aims to recognize hand gestures made by deaf and dumb individuals using real-time video processing and machine learning. Below Figure 1. represent the flow of proposed work for the sign language identification. First, we will capture live video using a camera or webcam, where the user performs hand gestures representing words in sign language. The video is continuously streamed and divided into individual frames in real time for processing. After that, each frame is resized and cropped to focus on the region of interest, typically the area containing the hand. This preprocessing step may also include background subtraction, contrast adjustment, and noise removal to enhance clarity and isolate the hand gesture more effectively.

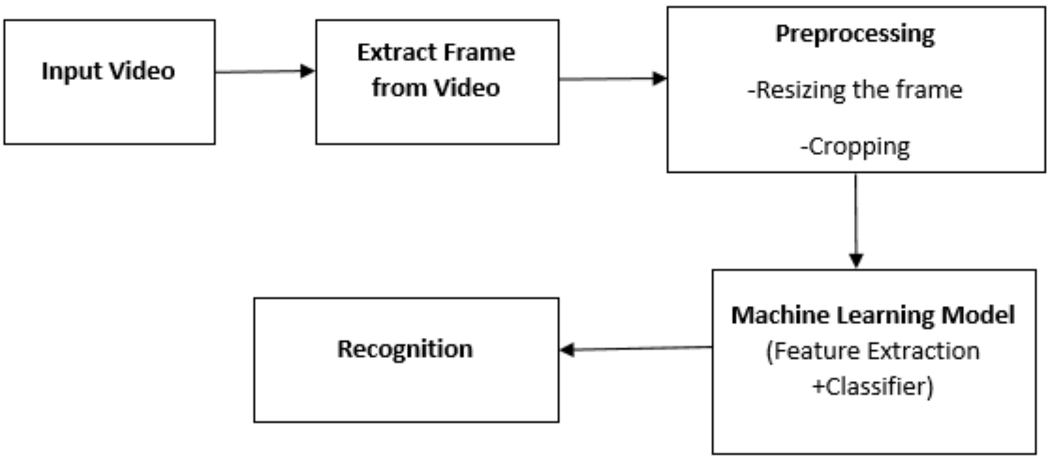

Fig 1. Flow of Proposed Work

Once the body is pre-processed, a system mastering or deep gaining knowledge of version which include a Convolutional Neural Network (CNN) [6,9,16,18]is implemented to extract vital spatial and temporal capabilities from the gesture. These features are then surpassed thru a skilled category model that identifies the precise gesture being proven. The classifier maps the gesture to its corresponding phrase or phrase from the educated dataset. Finally, the recognized word is displayed at the screen or converted to speech output. This actual-time method helps bridge communication among deaf and dumb individuals and the listening to network through translating sign language into understandable spoken or written phrases.

IV. Dataset

ASL Alphabet Dataset (Kaggle))

The ASL Alphabet Dataset available on Kaggle carries over 87,000 categorized RGB snap shots representing 26 hand signs of the American Sign Language (A–Z). Each photograph is 200x200 pixels and perfect for training CNN models for static gesture reputation. The dataset is user-pleasant and broadly adopted in research and training. It helps actual-time hand gesture class. It is freely reachable and appropriate for beginner to advanced ML initiatives.

RWTH-BOSTON-50 Dataset

This dataset incorporates continuous video sequences of 50 American Sign Language words, captured from multiple signers. It is designed for dynamic sign recognition and is commonly used in sequence modeling with RNNs or LSTM. Each video is labeled and preprocessed for gesture segmentation. It enables sentence-level translation research. Developed by RWTH Aachen University, it is available upon academic request.

Indian Sign Language (ISL) Dataset

This ISL dataset capabilities a group of static and dynamic hand signs and symptoms accomplished by means of more than one customers, protecting round 20 usually used Indian signs and symptoms. It incorporates each photographs and short movies. The dataset became created by using IIIT-Hyderabad and has been used in more than one Indian research papers. It helps real-time gesture recognition the use of CNN or hybrid models. Access is commonly granted for academic and studies purposes.

ASL Digits Dataset

This dataset consists of pix of hand gestures representing digits zero–9 in American Sign Language. It consists of over 2,000 grayscale snap shots with uniform background and lighting fixtures, making it perfect for digit classification the usage of easy CNN fashions. The dataset is light-weight, smooth to preprocess, and properly-suitable for real-time programs. It is beneficial for fundamental sign language systems which include numerical input or digit popularity. The dataset is open-source and to be had thru ML repositories and GitHub.

V. CHALLENGES

1. Variation in Hand Gestures: Different customers might also perform the identical gesture otherwise in phrases of pace, angle, and hand length, making it difficult for the version to generalize and preserve high accuracy throughout all customers.
2. Complex Backgrounds and Lighting Conditions: Real-time video may be captured in various environments, wherein inconsistent lighting fixtures and cluttered backgrounds can intervene with correct hand detection and gesture recognition.
3. Lack of Large and Diverse Datasets: High-best, categorized sign language datasets are restrained, especially for nearby or much less generally used sign languages. This shortage affects the training and overall performance of deep studying models.
4. Real-Time Processing Requirements: For powerful communication, the system have to apprehend gestures right away, which demands green algorithms and hardware able to processing video frames in real time without delay.
5. Gesture Overlap and Transition: Continuous gesture reputation is tough, as gestures may additionally circulate each other without clear barriers, making it tough to locate while one gesture ends and some other starts off.

CONCLUSION

The "Sign to Speech" machine helps deaf and mute people talk to others with the aid of changing hand signs into spoken words. It makes day by day communication less difficult and quicker without needing an interpreter or writing. The gadget works in actual time and uses neighborhood sign languages, so it feels natural for customers. It watches hand movements the usage of a camera and turns them into voice or text. This makes it useful in places like schools, hospitals, banks, and places of work. People with listening to or speech troubles can experience greater confident and independent. The machine also enables others apprehend and encompass them better.

REFERENCES

[1] S. Alyami, H. Luqman, M. Hammoudeh, Reviewing 25 years of continuous sign language recognition research: Advances, challenges, and prospects, Inf. Process. Manag. 61 (5) (2024) 103774.

[2] Y. Obi, K.S. Claudio, V.M. Budiman, S. Achmad, A. Kurniawan, Sign language recognition system for communicating to people with disabilities, Procedia Comput. Sci. 216 (2023) 13–20 7th International Conference on Computer Science and Computational Intelligence 2022.

[3] H. Alsolai, L. Alsolai, F.N. Al-Wesabi, M. Othman, M. Rizwanullah, A.A. Abdelmageed, Automated sign language detection and classification using reptile search algorithm with hybrid deep learning, Heliyon 10 (1) (2024) e23252.

[4] S. Arooj, S. Altaf, S. Ahmad, H. Mahmoud, A.S.N. Mohamed, Enhancing sign language recognition using CNN and sift: A case study on Pakistan sign language, J. King Saud Univ. Comput. Inf. Sci. 36 (2) (2024) 101934.

[5] Y. Liu, X. Jiang, X. Yu, H. Ye, C. Ma, W. Wang, Y. Hu, A wearable system for sign language recognition enabled by a convolutional neural network, Nano Energy 116 (2023) 108767.

[6] S.N. Koyineni, G.K. Sai, K. Anvesh, T. Anjali, Silent expressions unveiled: Deep learning for british and American sign language detection, Procedia Comput. Sci. 233 (2024) 269–278 5th International Conference on Innovative Data Communication Technologies and Application (ICIDCA 2024).

[7] J. Bora, S. Dehingia, A. Boruah, A.A. Chetia, D. Gogoi, Real-time Assamese sign language recognition using mediapipe and deep learning, Procedia Comput. Sci. 218 (2023) 1384–1393 International Conference on Machine Learning and Data Engineering.

[8] T. Liu, T. Tao, Y. Zhao, M. Li, J. Zhu, A signer-independent sign language recognition method for the single-frequency dataset, Neurocomputing 582 (2024) 127479.

[9] A Review Paper on Sign Language Recognition for The Deaf and Dumb published by R Rumana, Reddygari Sandhya Rani, and Mrs. R. Prema, in the year 2021.

[10] Indian Sign Language Recognition using Convolutional Neural Network published by Rachana Patil, Abhishek Bahuguna, and Mr. Gaurav Datkhile, in the year 2021.

[11] Machine Learning Techniques for Indian Sign Language Recognition published by Kusumika Krori Dutta, in the year 2017.

[12] Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique published by Joyeeta Singha, Karen Daas, in the year 2013.

[13] Indian Sign Artificial Neural Network Based Method for Indian Sign Language Recognition published by Adithya V., Vinod P. R., and Usha Gopalakrishnan, in the year 2013.

[14] Translation of Sign Language for Deaf and Dumb People published by Suthagar S., K. S. Tamilselvan, P. Balakumar, B. Rajalakshmi, C. Roshini, in the year 2020.

[15] Sign Language Recognition: State of The Art. ARPN Journal of Engineering and Applied Sciences Published by Sahoo, A. K., G. S. Mishra, and K. K. Ravulakollu in the year 2014.

[16] D.K. Singh, 3d-cnn based dynamic gesture recognition for indian sign language modeling, Procedia Comput. Sci. 189 (2021) 76–83 Computational Linguistics.

[17] Bhaidasna, Hetal, and Zubin Bhaidasna. "Object Detection Using Machine Learning: A Comprehensive Review." International Journal of Scientific Research in Computer Science, Engineering and Information Technology (2023): 248-255.

[18] Bhaidasna, Hetal, and Zubin Bhaidasna. “A Survey on Machine Learning Algorithms ", International Journal of Emerging Technologies and Innovative Research, ISSN:2349-5162, Vol.6, Issue 3, page no.53-56, March-2019.