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

Hand Speak's: Sign Language Recognition System

Siddhant Pandey[1], Aamir Khurshid[2], Shahbaz Ansari[3], Nagendra Nath Dubey[4]

[1,2,3,4]Dept. of Information Technology, ABES Institute of Technology GHAZIABAD, INDIA

Abstract
HAND SPEAK'S is a sign language recognition system. The system is designed to recognize and interpret a wide range of sign language gestures, converting them into readable text or spoken words in real-time. The sign language recognition system presented in this paper represents a significant step towards improving accessibility and inclusivity for the deaf and hard-of-hearing communities. The system integrates a user-friendly interface that allows users to interact with the recognition software seamlessly. This research paper presents a novel approach to recognize American Sign Language (ASL) at the sentence level utilizing Convolutional Neural Networks (CNN) with TensorFlow and OpenCV. In this paper, we explore the challenges associated with ASL recognition, including the complexities of hand gestures, spatial variations, and real-time processing requirements. The proposed method aims to enhance the accuracy and efficiency of ASL recognition systems, thereby facilitating smoother interaction and understanding between individuals using sign language and those who do not. The study discusses the dataset preparation, model architecture, training process, and evaluation metrics. Experimental results demonstrate the effectiveness and robustness of the proposed approach in recognizing ASL sentences accurately. Our method provides 97.2 % accuracy. Future work will focus on expanding the enhancing real-time performance, and exploring multilingual sign language recognition capabilities.

Keywords: American Sign Language, CNN, TensorFlow, OpenCV, Gesture Recognition, Sentence Level Recognition.

1. INTRODUCTION

Hand gesture recognition is also in Human-Computer Interaction (HCI) because it interacts with the user directly. Human-Computer Interaction (HCI) is the study, planning, or design of interaction between users and computers(Obi et al., 2022). Sign language plays an important role for communication between individuals with hearing and speech impairments, particularly for the Dumb and Deaf community. The ability to express thoughts, emotions, and ideas through gestures plays a fundamental role in bridging the communication gap for those who cannot rely on spoken language. American Sign Language (ASL) stands out as a predominant form of sign language, offering a visual means of communication for Dumb and Deaf (D&M) individuals.

Sign language is a visual means of communication that uses hand gestures, facial expressions, and body movements to convey meaning. It serves as the primary mode of communication for many deaf and hard-of-hearing individuals worldwide, enabling them to interact with others and express themselves effectively. The advances in machine learning and deep learning technology are providing new methods and algorithms for recognizing Indian sign language alphabets efficiently, accurately and inexpensively(Katoch et al., 2022). The end-to-end auto run of these models overcomes the highly subjective and inconsistent limitations of traditional methods, improving accuracy and efficiency of the results.

Sign language can be divided into two main categories: static and dynamic. Static signs are steady hand and face gestures, while dynamic signs are further divided into isolated signs and continuous signs. (Kothadiya et al., 2022) Isolated signs are hand gestures and facial expressions for a single word, i.e., 'home', 'love', and many more, while continuous signs are a sequence of isolated signs involving both hand gestures and facial expressions, i.e., 'welcome to my new home'.For D&M individuals, whose only disability lies in communication, sign language emerges as the primary avenue for interaction.

It is no doubt that communication plays a vital role in human life(Godage et al., 2021). There is, however, a significant population of hearing-impaired people who use non-verbal techniques for communication, which a majority of the people cannot understand. The predominant of these techniques is based on sign language, the main communication protocol among hearing impaired people.

Sign languages need an intelligent system to translate sign language to another based on natural languages. It is hard for most people who are not interested in sign language(Elsayed & Fathy, 2020). These nonverbal exchanges, known as gestures, are comprehended through visual perception, forming the basis of sign language.

Human activity recognition is an important and difficult topic to study because (Hernandez et al., 2020) of the important variability between tasks repeated several times by a subject and between subjects.

By utilizing their hands to convey different gestures and expressions, D&M individuals can effectively communicate their ideas and feelings to others.

Since the sign language has become a potential communicating language for the people who are deaf and mute, (Kamruzzaman,2020) it is possible to develop an automated system for them to communicate with people who are not deaf and mute.

Motion of any body part like face, hand is a form of gesture. Gestures are performed by deaf and dumb community to perform sign language. Sign language can be performed by using Hand gesture either by one hand or two hands.

Deafness and vocal disability bring them significant communication problems while accessing education, job, etc. Since the majority of normal people can't understand their language, communication using sign language is always limited in the deaf-dump community. As humans, they deserve to get all the help needed to live an ordinary life. One way to help them is by using advanced technology to overcome some of the difficulties they face. Sign language using hand gestures are helpful for establishing human-machine interaction, which can facilitate the communication between normal people and hard of hearing people with the machine as a mediator.

It is of two type Isolated sign language and continuous sign language. Isolated sign language consists of single (Athira et al., 2022) gesture having single word while continuous ISL or Continuous Sign language is a sequence of gestures that generate a meaningful sentence.

Sign language (SL) is a natural, visual, and non-verbal Language(Hayani et al., 2019). It's a rich and expressive form of communication that enables deaf individuals to fully participate in society and express themselves. Some countries such as Belgium, the UK, the USA or India may have more than

one sign language. Many sign languages are in used all over the world, for instance, Indian Sign Language, British Sign Language (BSL).
Sign language is a visual language and consists of 2 major components:

Table 1.1

Fingerspelling	Word level Sign vocabulary
Used to spell words letter by letter.	Used for the major part of communication.

It is commonly taught in schools for the deaf and may also be learned by hearing individuals who wish to communicate with deaf or hard-of-hearing individuals. Sign language interpreters play a crucial role in facilitating communication between deaf and hearing individuals in diverse settings, including conferences, meetings, and public events.
Sign languages vary by region and country, just like spoken languages, and they have their own grammar and vocabulary. In addition to manual signs, sign languages often incorporate facial expressions and body postures to convey nuances of meaning and emotion.

Fig 1.1 Cropped image of montage panel of various users and backgrounds for American Sign Language Letters from sign language MNIST

2. RELATED WORK

This research employs Convolutional Neural Networks (CNN) to classify hand gestures from American

Sign Language (ASL) datasets. [1] Initially, each hand image undergoes processing through a specialized filter. The primary focus of this study is to enhance the accuracy of gesture recognition. The system demonstrated a remarkable accuracy rate of 96.3% across the 26 letters of the alphabet.

In this study, we introduce a method leveraging the Bag of Visual Words (BOVW) model for recognizing the alphabets (A-Z) and digits (0-9) of Indian Sign Language in real-time video streams. The system outputs the predicted labels both textually and through synthesized speech. The segmentation process utilizes skin color detection and background subtraction techniques to isolate the gestures for analysis. [2] Feature extraction is performed using Speeded Up Robust Features (SURF), creating histograms that associate specific gestures with their corresponding labels. For classification, the model integrates both Support Vector Machines (SVM) and Convolutional Neural Networks (CNN), aiming to enhance the accuracy and robustness of the recognition process.

This paper proposes a deep learning-based approach for detecting and recognizing words conveyed through gestures, focusing on Indian Sign Language (ISL). [3] We utilize feedback-based learning models, specifically Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), to analyze and interpret signs from isolated video frames of ISL. Our exploration includes testing four sequential configurations of these models, integrating two layers each of LSTM and GRU in various combinations. Among these, the proposed model architecture, which arranges a single layer of LSTM followed by a layer of GRU, demonstrates a robust performance, achieving approximately 97% accuracy across 11 distinct signs. This configuration optimally harnesses the strengths of both LSTM and GRU, leading to high precision in gesture recognition.

Our proposed system allows users to sign full sentences seamlessly using two Myo armbands that capture gestural (EMG) and spatial (IMU) data. [4] We trained our model on a dataset of 49 words and 346 sentences from a single signer, achieving 75-80% word-level accuracy and 45-50% sentence-level accuracy. This setup highlights our system's capability to effectively process and recognize continuous sign language gestures.

This study proposes a method to classify 60 signs from American Sign Language using data captured by the Leap Motion sensor. [5] We explored various machine learning and deep learning models, including a hybrid model known as DeepConvLSTM, which combines convolutional layers with recurrent Long-Short Term Memory (LSTM) cells. Our findings indicate that the DeepConvLSTM model and a conventional convolutional neural network (CNN) achieved the highest accuracies among the tested models, recording 91.1% (±3.8) and 89.3% (±4.0) respectively. These results underscore the effectiveness of integrating convolutional and recurrent architectures for enhancing sign language recognition accuracy.

The proposed system is designed to automatically detect hand sign letters and verbally articulate the results in Arabic using a deep learning model. [6] With a recognition accuracy of 90% for Arabic hand sign-based letters, this system proves to be highly reliable and effective for users.

This paper introduces a novel system leveraging Convolutional Neural Networks (CNNs) to automatically recognize numbers and letters in Arabic sign language using real-world data. The study includes a comparative analysis, demonstrating the effectiveness and robustness of the proposed CNN-based approach against traditional methods such as k-nearest neighbors (KNN) and support vector machines (SVM). CNNs have transformed Sign Language Recognition (SLR) by autonomously extracting features from image data, resulting in notable accuracy enhancements. [7] Nevertheless, their computational demands and reliance on extensive datasets for optimal performance are noteworthy considerations. Despite these challenges, the CNN-based system in this study showcases promising advancements in Arabic sign language recognition.

The proposed model utilizes Support Vector Machines (SVM) with a dataset consisting of 900 static images and 700 videos to evaluate alphabets in sign language. [8] It achieves an accuracy rate of 91% for finger-spelling alphabets and 89% accuracy for recognizing single-hand dynamic words. These results demonstrate the effectiveness of SVM in accurately classifying sign language gestures.

Table.1. Summary of significant work on the basis of evaluation metrics

S. no.	Reference No.	Proposed Work	Evaluation Matrix
1.	(Obi et al., 2022)	The proposed model is created using the two layers Convolutional Neural Network (CNN) with 30526 images.	Accuracy rate of 96.3%.
2.	(Katoch et al., 2022)	SURF (Speeded Up Robust Features) features have been extracted from the images and histograms are generated to map the signs with corresponding labels. The proposed model uses SVM and CNN for classification. 36000 images used for alphabets and digits.	SVM has given an accuracy of 99.14% and CNN has given an accuracy of 94%
3.	(Kothadiya et al., 2022)	The proposed model, consists of a single layer of LSTM followed by GRU and works for 11 words, for each word about 1100 video sample.	Accuracy rate of around 97% was achieved with 80% training data.
4.	(Godage et al., 2021)	The system employs two Myo armbands for capturing gestures. Through signal processing and supervised learning techniques, it interprets data collected from these armbands, which includes a vocabulary of 49 words and 346 sentences.	Word level accuracy of 75-80%.
5.	(Hernandez et al., 2020)	The proposed model uses DeepConvLSTM that integrates convolutional and recurrent layers with LSTM cells. The model uses 16890 labeled signs	Accuracy of 91.1% using leap motion.
6.	(Kamruzzaman, 2020)	The proposed model uses CNN with 3875 images. For every letter 100 images used for learning and 25 images for testing.	Accuracy of 90.02% was achieved with 80% training data. It also converts text to speech.
7.	(Hayani et al., 2019)	The proposed model uses CNN with 2030 images of numbers and 5839 images of	Accuracy of 90.02% was achieved with

		letters.	80% training data.
8.	(Athira et al., 2022)	The proposed model uses SVM with 900 static images and 700 videos to test alphabets.	Accuracy rate of 91% for finger-spelling alphabets and 89% accuracy of single hand dynamic word.

3. PROPOSED WORK

The development of a vision-based sign language recognition system for Dumb and Deaf (D&M) individuals involves a systematic approach that encompasses data collection, preprocessing, feature extraction, and classification algorithms. By integrating advanced technologies and machine learning techniques, researchers aim to create a robust and accurate system capable of interpreting sign language gestures in real-time.

Image acquisition: Image acquisition involves the utilization of a variety of cameras, illumination sources, and equipment to acquire the necessary data and high-resolution images essential for a particular application.

Data Collection: The first step in the methodology involves collecting a diverse dataset of sign language gestures, including the ASL alphabet and common hand movements used in communication. This dataset serves as the foundation for training and testing the sign language recognition system, ensuring that it can accurately interpret a wide range of gestures and symbols. Data collection may involve recording videos of sign language gestures performed by individuals proficient in ASL or other sign languages, capturing variations in hand shapes, movements, and expressions.

Preprocessing: Once the dataset is collected, preprocessing techniques are applied to enhance the quality and usability of the data. Preprocessing steps may include noise removal, image enhancement, normalization, and segmentation to prepare the input data for feature extraction and classification. By cleaning and optimizing the dataset, researchers can improve the performance and accuracy of the sign language recognition system, reducing errors and enhancing the overall user experience.

Segmentation: Segmentation in hand detection involves separating the hand region from the background or other objects in an image or video frame. Hand tracking algorithms can maintain continuity in hand position across frames in a video sequence. Effective segmentation is crucial for accurate hand detection and training systems used in applications such as sign language recognition and gesture control.

Feature Extraction: Feature extraction plays a crucial role in identifying key characteristics and patterns within the sign language gestures, enabling the system to differentiate between different symbols and gestures. Common features extracted from the hand movements may include hand shape, finger positions, movement trajectories, and temporal dynamics. By capturing these essential features, researchers can develop robust algorithms for gesture classification and interpretation, facilitating accurate and reliable recognition of sign language gestures.

Classification Algorithms: The final step in the methodology involves implementing classification algorithms to interpret and translate the extracted features into meaningful symbols or text. Algorithms are trained on the preprocessed data to classify and recognize sign language gestures in real-time. By leveraging advanced classification techniques, researchers can achieve high accuracy and efficiency in interpreting sign language gestures, enabling seamless communication for D&M individuals.

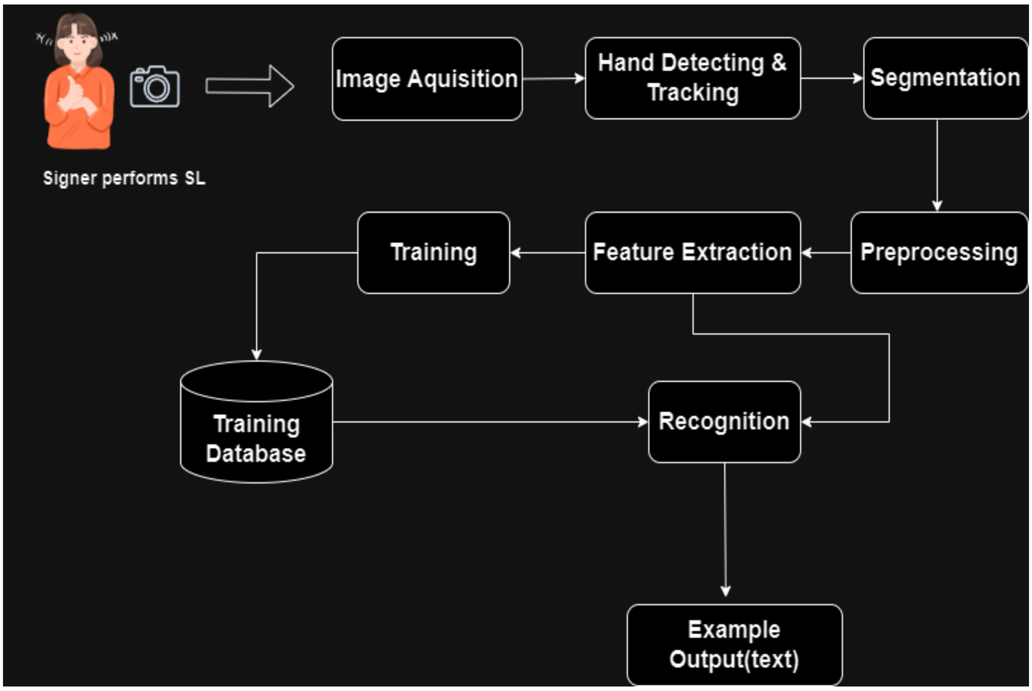

Fig.3.1 Sign Language Recognition System Diagram

4. Implementation

Step 1. Data collection

1. Initialize the camera and necessary modules (e.g., hand detection, classifier).
2. Start the main loop:
- Read a frame from the camera.
- Detect hands in the frame.
- If hands are detected:
- Select the first detected hand.
- Extract the bounding box coordinates of the hand.
- Create a white background image of predefined size.
- Crop the hand region from the frame.
- Calculate the aspect ratio of the cropped region.
- Resize and place the cropped region onto the white background.
- Display the cropped hand region and the white background image.
- Display the original frame.
3. Wait for user input:
- If the user presses 's':
- Save the white background image to a file with a timestamp.

Step 2. Hand Detection:

1. For each detected hand:
- Get the bounding box coordinates of the hand.
- Draw a filled rectangle as background for the text label.
- Put the text label on the image.

- Draw a rectangle around the detected hand.
2. Display the cropped hand region and the white background image.
3. Display the original frame with annotations.

Step 3. Sign Recognition:
1. Create a white background image of specified size.
2. Crop the hand region from the frame.
3. Calculate the aspect ratio of the cropped hand region.
4. If the aspect ratio is greater than 1:
- Resize the cropped image to fit into the white background.
- Compute the prediction and index using the classifier.
- Print the prediction and index.
5. Else:
- Resize the cropped image to fit into the white background.
- Compute the prediction and index using the classifier.
- Print the prediction and index.

 # End of the Loop:
- Continue the loop until manually stopped.

5. Result:
To improve the proposed system, we produced a model with training and validation accuracy of 97.29% and 99.6%, respectively. We have visualized the accuracy, and the loss from our training and validation model of our research algorithm in the form of graphs.

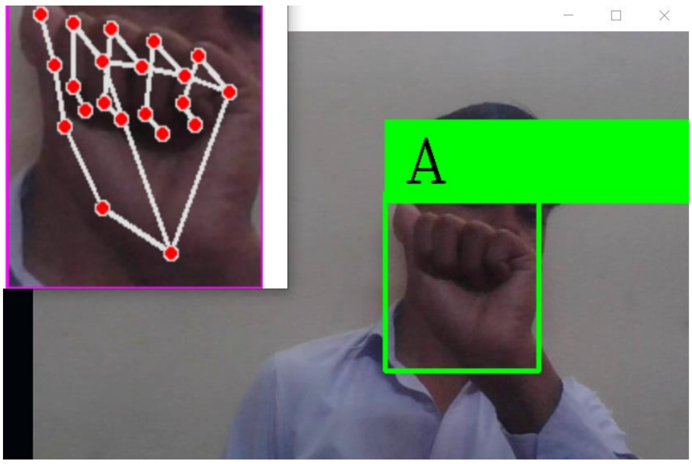

Fig 5.1. Sign detected "A"

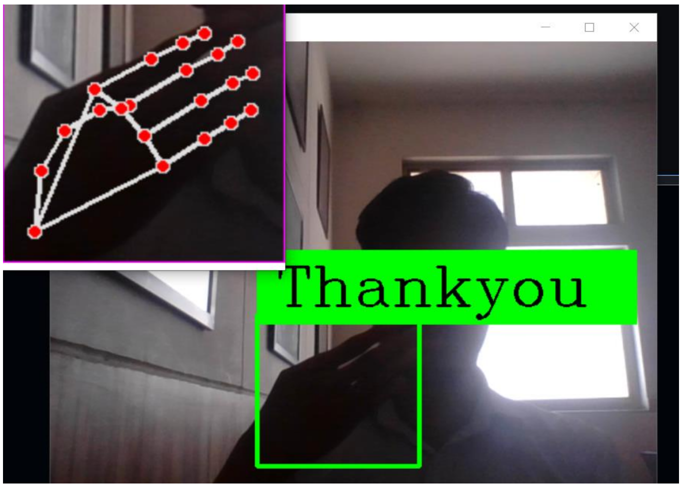

Fig 5.1. Sign detected "Thankyou"

Table 2.: Sign Wise Classification Accuracy

Sign	Recognition Accuracy	Average Accuracy
A	88.49%	97.29%
B	100%	
C	91.67%	
D	99.60%	
E	81.35%	
F	98.02%	
G	99.21%	
H	98.81%	
I	99.60%	
J	95.63%	
K	92.46%	
L	100%	
M	87.70%	
N	86.90%	
O	82.94%	
P	92.46%	
Q	99.60%	
R	96.03%	
S	98.81%	
T	94.44%	
U	94.05%	
V	96.83%	
W	97.62%	
X	96.43%	
Y	98.02%	

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Z	92.06%	

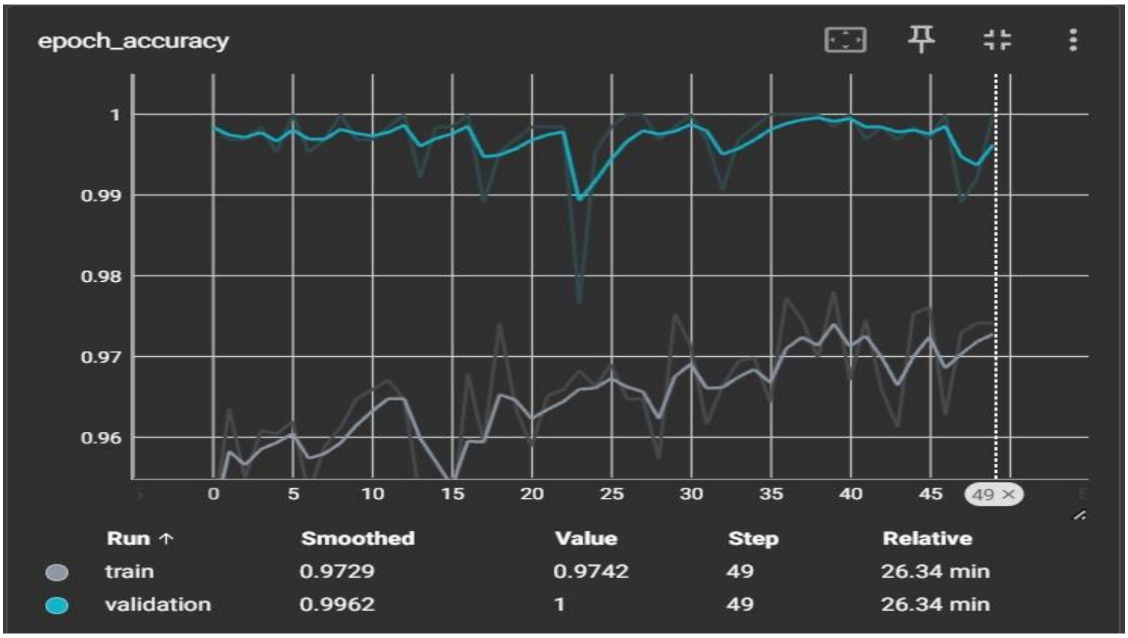

Fig. 5.2. Epoch Accuracy graph

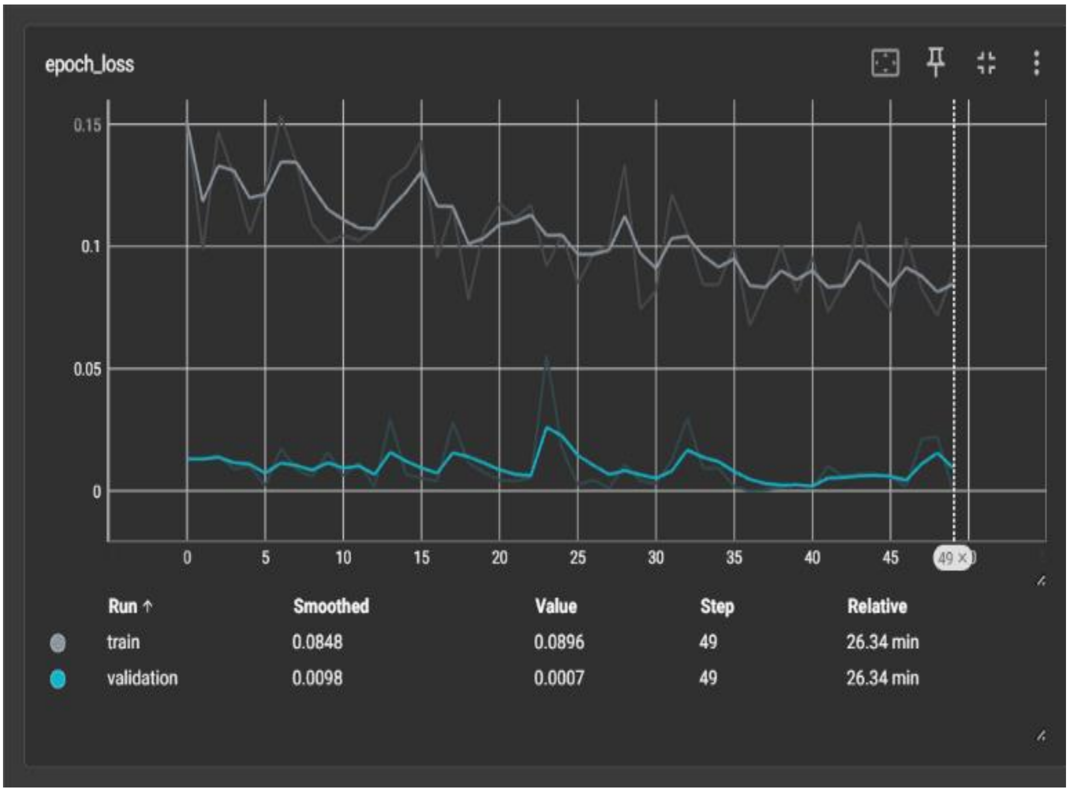

Fig. 5.3. Epoch Loss graph

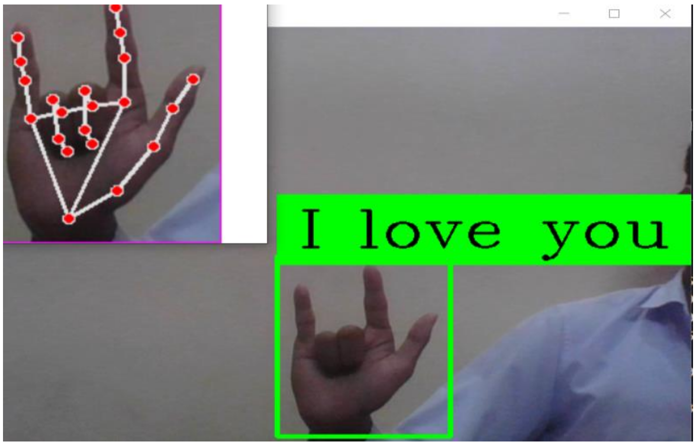

Fig 5.2. Sign detected "I Love You"

6. Conclusion:

A significant challenge in real-life applications is hand gesture recognition in terms of the accuracy and robustness associated with it. Non-touch hand gesture recognition of ASL is presented in this paper using the webcam.

Our research endeavors in enhancing the proposed Sign Language Recognition (SLR) system have yielded promising results, as evidenced by the significant improvements in training and validation accuracies. Through various experiments involving adjustments to

the number of training and testing sets, we have fine-tuned our model to achieve commendable accuracy rates.

The overall classification accuracy of 97.29% signifies the system's ability to reliably interpret a

wide range of ASL signs, thereby facilitating effective communication for individuals with disabilities who rely on ASL as their primary means of expression.

In conclusion, our research represents a significant contribution to the field of Human-Computer Interaction, particularly in the realm of assistive technology for individuals with disabilities.

7. Future Scope:

This project can be enhanced by being built as a web/mobile application for the users to conveniently access the project. Also, the existing project only works for ASL; it can be extended to work for other native sign languages with the right

amount of data set and training. This project implements a finger spelling translator; however, sign languages are also spoken in a contextual basis where each gesture could represent an object, or verb. So, identifying this kind of a contextual signing would require a higher degree of processing and natural language processing (NLP).

Sign language recognition systems can be integrated into existing communication tools and platforms to enhance accessibility for deaf and hard-of-hearing individuals. For example, video conferencing software could incorporate real-time sign language interpretation features, enabling deaf users to participate in meetings and discussions more seamlessly.

8. Reference

1. Obi, Y., Claudio, K. S., Budiman, V. M., Achmad, S., & Kurniawan, A. (2022). Sign language recognition system for communicating to people with disabilities. *Procedia Computer Science*, *216*,

13–20. https://doi.org/10.1016/j.procs.2022.12.106
2. Katoch, S., Singh, V., & Tiwary, U. S. (2022). Indian Sign Language recognition system using SURF with SVM and CNN. *Array*, *14*. https://doi.org/10.1016/j.array.2022.100141
3. Kothadiya, D., Bhatt, C., Sapariya, K., Patel, K., Gil-González, A. B., & Corchado, J. M. (2022). Deepsign: Sign Language Detection and Recognition Using Deep Learning. *Electronics (Switzerland)*, *11*(11). https://doi.org/10.3390/electronics11111780
4. Godage, I., Weerasignhe, R., & Sandaruwan, D. (2021). *Sign Language Recognition for Sentence Level Continuous Signings*. 53–73. https://doi.org/10.5121/csit.2021.112305
5. Hernandez, V., Suzuki, T., & Venture, G. (2020). Convolutional and recurrent neural network for human activity recognition: Application on American sign language. *PLoS ONE*, *15*(2). https://doi.org/10.1371/journal.pone.0228869
6. Kamruzzaman, M. M. (2020). Arabic Sign Language Recognition and Generating Arabic Speech Using Convolutional Neural Network. *Wireless Communications and Mobile Computing*, *2020*. https://doi.org/10.1155/2020/3685614
7. Hayani, S., Benaddy, M., El Meslouhi, O., & Kardouchi, M. (2019, July 1). Arab Sign language Recognition with Convolutional Neural Networks. *Proceedings of 2019 International Conference of Computer Science and Renewable Energies, ICCSRE 2019*. https://doi.org/10.1109/ICCSRE.2019.8807586
8. Athira, P. K., Sruthi, C. J., & Lijiya, A. (2022). A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario. *Journal of King Saud University - Computer and Information Sciences*, *34*(3), 771–781. https://doi.org/10.1016/j.jksuci.2019.05.002
9. Boháček, M., & Hrúz, M. (n.d.). *Sign Pose-based Transformer for Word-level Sign Language Recognition*. https://developer.apple.com/documentation/
10. Islam, M. R., Mitu, U. K., Bhuiyan, R. A., & Shin, J. (2018). Hand gesture feature extraction using deep convolutional neural network for recognizing American sign language. *2018 4th International Conference on Frontiers of Signal Processing, ICFSP 2018*, 115–119. https://doi.org/10.1109/ICFSP.2018.8552044
11. Koppalkar, P., J, B. M., Bhaskar, H., & Hullali, A. (n.d.). *Indian Journal of Software Engineering and Project Management (IJSEPM) Sign Language Recognition System*. https://doi.org/10.54105/ijsepm.C9011.011322
12. Buttar, A. M., Ahmad, U., Gumaei, A. H., Assiri, A., Akbar, M. A., & Alkhamees, B. F. (2023). Deep Learning in Sign Language Recognition: A Hybrid Approach for the Recognition of Static and Dynamic Signs. *Mathematics*, *11*(17). https://doi.org/10.3390/math11173729
13. Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M. A., Alrayes, T. S., Mathkour, H., & Mekhtiche, M. A. (2020). Deep learning-based approach for sign language gesture recognition with efficient hand gesture representation. *IEEE Access*, *8*, 192527–192542. https://doi.org/10.1109/ACCESS.2020.3032140
14. Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M. A., & Mekhtiche, M. A. (2020). Hand Gesture Recognition for Sign Language Using 3DCNN. *IEEE Access*, *8*, 79491–79509. https://doi.org/10.1109/ACCESS.2020.2990434
15. Aly, S., & Aly, W. (2020). DeepArSLR: A novel signer-independent deep learning framework for isolated arabic sign language gestures recognition. *IEEE Access*, *8*, 83199–83212.

https://doi.org/10.1109/ACCESS.2020.2990699

16. Goyal, A., Dhar, D., Nair, P. A., Saini, C., & Prakash, C. (n.d.). *Sign Language Recognition System*. https://github.com/shackenberg/Minimal-Bag-of-Visual-Words-Image-
17. Gunji, B. M., Bhargav, N. M., Dey, A., Zeeshan Mohammed, I. K., & Sathyajith, S. (2022). Recognition of Sign Language Based on Hand Gestures. *Journal of Advances in Applied & Computational Mathematics*, *8*, 21–32. https://doi.org/10.15377/2409-5761.2021.08.3
18. Sarkar, K., Shivramwar, C., & Mohabe, D. (2023). Sign Language Interpreter. *International Journal for Research in Applied Science and Engineering Technology*, *11*(4), 2253–2257. https://doi.org/10.22214/ijraset.2023.50527