Deep Learning-Driven Character Classification in Indian Sign Language: A Real-Time Approach

[1] Neha Ranjan, [2]Shalini Mhaske, [3]Omkar Pattnaik, [4]Prajakta Dawange, [5]Gitanjali Khatik
[1,2,4,5] *Student, School of Computer Science and Engineering, Sandip University, Nashik, Maharashtra-422213, India.*
[3]*Associate Professor, School of Computer Science and Engineering, Sandip University, Nashik, Maharashtra-422213, India.*

Abstract- **It can be quite difficult to communicate with someone who has a hearing loss. The study in this paper aims to investigate the challenges associated with character classification in Indian Sign Language (ISL). Sign dialect communication is ineffective for people with speech inadequacies or hearing impairments. To someone who has never learned this lingo, the signals given by those who are inadequate are confused or jumbled. It is necessary to communicate in both directions. In this work, we present an affirmation in Sign Dialect utilizing Indian Sign Lingo. The client must be able to record hand movements with a webcam for this test, and the system must be able to identify and show the picture's title. The obtained image is processed through a number of planning procedures that incorporate several computer vision methods, such as grayscale conversion, cover operation, and extension. Convolutional Neural Networks (CNN) are utilized for image recognition and display organization. The final perfection of our appearance is 95%.**

***Keyword* - Convolutional Neural Network (CNN), Hand Gesture, Sign Language, Sign Language Recognition (SLR).**

1. INTRODUCTION

Information sharing is one of the most important things that people need to do in order to live in society. Hearing-impaired persons interact with each other via sign language, which is difficult for non-hearing-impaired people to grasp. Despite the fact that American Sign Dialect has not been extensively studied, Indian Sign Dialect is very different from "American Sign Dialect". ISL uses two hands (20 of 26 hands) for communication, whereas ASL uses one hand. When using both hands, the highlights are often obscured because both hands cover. ISL gesture research is also limited by the requirement for documentation and local sign dialect contrasts. This study highlights the initial phase of employing "Indian Sign Dialect" to facilitate communication between hearing and hearing-impaired individuals. This stretches into a single sentence and a common term rather than being motivating. Hearing helped people learn the skills they needed to enhance their communication skills, but it also hindered their capacity to interact with the outside world. This document aims to help you learn the Indian Sign Dialect by heart. Collection of letters using hand motion. Although it has received little attention in ISL, comprehending motions and sign dialect could be a well-researched issue in American Sign Dialect. This problem needs to be clarified, but rather than employing state-of-the-art tools like gloves or Kinect, we need to study the signals from images (which are viewable through a webcam) and then utilize computer vision and machine learning (ML) to separate and arrange of the greatest interest parts.

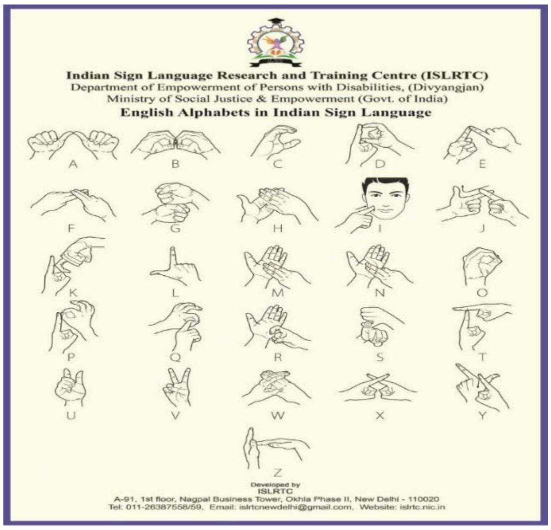


Figure 1: Indian Sign Language Alphabets.

1.1 Problem Statement

The goal of the "Sign Language Recognition System" is to develop a reliable, accurate, and effective system for deciphering the meaning of hearing-impaired and dumb signals and translating them into easily understood text. This strategy aims to enhance communication between deaf and hard of hearing people and sign language beginners.

2. LITERATURE SURVEY

People who suffer from hearing loss may become estranged and desolate, which can significantly affect their social and professional lives. Hearing people communicated with each other using sign dialect. When a regular person has to converse with someone who has hearing impairments, a translator is necessary. To deal with persons who are hard of hearing, many people never attempt to learn sign language. Finding a skilled and knowledgeable interpreter for sign dialect becomes difficult every time and everywhere. According to S. Reshna's presentation, sign language robotization. The recognition framework has been studied for many years and could be a step toward improving the lives of the hearing-impaired community. [1] [5] Rajat Soni has suggested a motion affirmation method for communication. This method's two guiding steps are specifically designed to safeguard data, prepare the final CNN classifier, and test it. The standard system stream chart addressing the operational display of the system is depicted in the accompanying fig. The next stage is to plan and test the CNN classifier: The intended demonstration will comprise the NVIDIA GeForce GTX 1650 “Graphical Preparing Unit (GPU)”, 1000GB of solid-state drive (SSD), 4 GB of Slam, and 16GB of irregular access memory (Slam). About 30 different models were made using a variety of convolutional, pooling, smooth, dropout (to prevent overfitting of the demonstration), and thick layers. These models were then adjusted using the Keras tuner to obtain the optimal hyperparameter settings. [2] Ankita Wadhawan and her group are focusing on information security and data preparation. Information security's picture preprocessing module receives three-channel RGB picture outlines that are taken from the camera.

The dataset is a collection of RGB images for several types of inactive indicators. Each of the inactive signs has 350 photos out of the dataset's 35,000 total. Among the 100 distinct sign classes are ten numerals, twenty-three sets of English letters, and sixty-seven frequently used phrases (such as stand, bowl, fever, water, hand, etc.). The dataset contains inactive sign images of different sizes, colors, as well as natural environments to help the classifier's better generalization. In information preprocessing, a number of morphological procedures have been employed to eliminate noise from data. In this step, the sign images have been prepared by employing 2 preprocessing techniques: picture resizing and normalization. A photo becomes 128 when it is scaled.

Now that the pixel intensity values in these images have been standardized, variance and cruel 0 result.[3] The study's main goals were to produce a lot of data for Indian Sign Dialect, design the optimal CNN presentation for the preprocessed image, and attain the highest accuracy. [4] Using the proper gesture, Rachana Patil aimed to view letter sets in Indian Sign Dialect.

Unlike American Sign Dialect, where the distinction between motions and sign dialects was extensively researched, Indian Sign Dialect has paid less attention to this issue. The research's objective was to comprehend this issue, but rather than employing pricey tools like gloves or the Kinect, they needed to be able to identify motions in pictures that could be viewed through a webcam. After that, they classified and identified certain highlights using machine learning and computer vision algorithms.[6] Based on Chandra Mani Sharma and her team, people with speech and hearing impairments can communicate thanks to advanced technology. Engage in conversation with others. One such range that has piqued the curiosity of scholars across several fields is automated sign dialect recognition. This work separates ISL basic graphics from 35 different classes by employing a DL method based on computer vision. The demonstration has exceptional misfortune and accuracy throughout preparation, and it can achieve 100% correctness on undetected test data. For those who have hearing or speech problems, it might be a useful tool for communication. [8] Some examples of different sign dialects are ISL, “British Sign Dialect (BSL)”, and “American Sign Language (ASL)”. The vocabulary of sign dialects is similar to that of spoken languages, which have a vocabulary of words. Sign dialects are employed differently in different countries and are neither standardized nor consistent. A manual sign dialect translator frequently infringes on the subject's right to security and isn't always a smart option. Robotized sign dialect interpreters that convert sign language into spoken or written language solve this difficulty.[9] Pias Paul used the “ASL Finger Spelling Dataset” and several CNN-based algorithms to classify 24 characters. They showcased exchange-learning strategies and models designed specifically to tackle this problem. In contrast to the best distributed result that is currently in use, one of their own models got a test precision of 86.52%. [10]

3. PROPOSED METHODOLOGY

The continuous ISL dataset is used to conduct experiments. Different lighting and backdrop conditions are used in its creation, along with changes to the surroundings, attire, and gestures. The hand image's distinguishing characteristics are taken out and categorized in order to use CNN to identify the sign.[1][2][14]

3.1 Image Acquisition

The camera's color images are transmitted to the image pre-processing module using the data acquisition application we developed. A collection of RGB photos with different static characters makes up the dataset. There are 74,188 static character images in the dataset. There are 33 distinct categories of people using 23 sets of English letters and numbers from 0 to 9. [39] To improve the classifier's accuracy, the dataset includes static sign images in a range of sizes, colors, and environmental conditions. [2] [3] [21]

3.2 Preprocessing

We gathered our original, 128-resolution photos.
This helps us avoid higher resolution data and save storage. Adding a Gaussian blur to the grayscale image is one of the first preprocessing processes that helps us eliminate extraneous noise and high-frequency areas of the image. Our subsequent actions involve applying thresholding, which transforms our blurred grayscale photos into binary inverted images. Here, we employ Otsu's approach since it provides the ideal thresholding value we require and aids in the division of our photos into foreground and background. Another version of our dataset was used to test several techniques, such as contours for manual segmentation and preprocessing, which were eliminated from the final preprocessed dataset since they didn't meet our needs.

Following thresholding, our dataset now includes polarizing photos, each of which has a segmented hand because the hand portion of the image is white. To help construct the best hand segments for our dataset, the lighting had to be the most balanced of all the characteristics. We used a variety of data augmentation methods, including zooming, shearing, rotation, and horizontal flipping, on our dataset. We used an interval between each class's photographs to execute augmentation for a certain number of them, producing several variants of those images that were subsequently merged with our original dataset. Because enhancing every image would make our dataset too large to handle, given our constraints, the interval was added. For every augmentation step, we have used the Image Data Generator package in Python, which is offered by Keras. On a different copy of our dataset, we also tested with additional methods like "shift range" and "fill mode," but this led to undue variety within a class, which is not what we wanted.

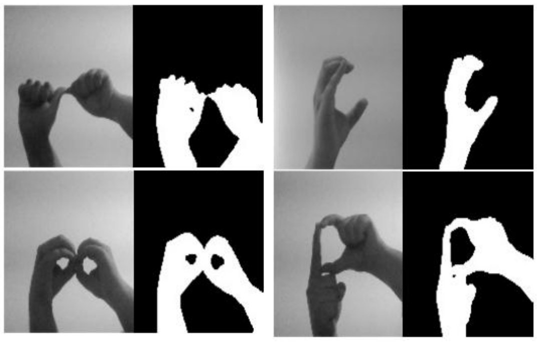


Figure 2: Original Hand Gesture Recognition

3.3 Recognition

Our final output and forecast are made in real time after our model has been successfully created and trained. In front of the device camera, the ISL signs are displayed. The live video frames are then taken from the feed, and each frame is processed by applying thresholding and blurring, just like the training dataset. Next, using its training data, the CNN model predicts the performed sign. The anticipated label and accuracy % are then shown. The finished user interface enables the user to view the expected alphabet and the alphabet to build words by pressing a button. Furthermore, users can create their own dataset and train the model with distinctive sign language data through the provided GUI. To generate a hand segment during real-time prediction, the user must ensure sufficient lighting and modify the background to a uniform color. We additionally display the threshold view, enabling the user to regulate the environment according to the binary feed. We employ the Python tkinter package to construct our interface. Furthermore, we include a text-to-speech feature that enables the projected alphabet to be articulated audibly for the user. If students wish to learn how to create the signs, they may also review the sign language images corresponding to each letter of the alphabet. Enhanced user communication is facilitated by these elements in our interface.

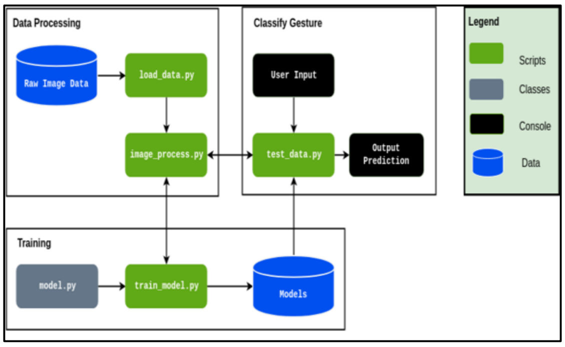


Figure 3: Proposed Methodology

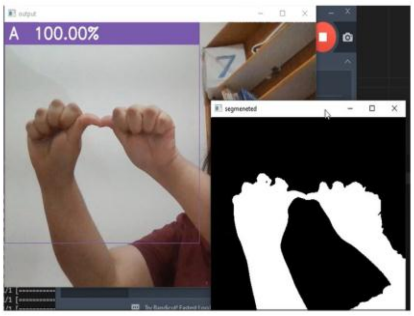
3.4 Model Building and Training

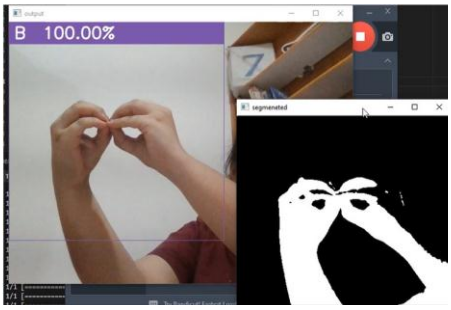
CNNs are often utilized for image classification due to their proficiency in finding, extracting, and categorizing essential features from images. To evaluate the CNN model, we extract and classify features. Convolutional Neural Networks extract features well. As the neural network and feature extraction increase in complexity or layer count, CNNs often exhibit enhanced picture recognition accuracy. Consequently, we utilize multiple layers in our model. Our initial ISL dataset had 13,000 photographs, and a simple two-layered CNN model was employed, yielding satisfactory results. As our dataset expanded, it became necessary to increase the model's complexity to achieve satisfactory results for the larger dataset. The initial three convolutional layers of our CNN model did not yield optimal results for real-time predictions in the latest iteration of our dataset, comprising 40,000 images. Consequently, we augmented the batch size to 128. Upon evaluating multiple iterations of the model, we finalized a CNN model for our experiment of 3 dense layers and four convolutional layers. A max pooling layer and a dropout layer come after the first convolutional layer, which has 32 filters and a 3x3 kernel size. Our future layout consists of a max pooling and dropout layer after a convolutional layer with 64 filters. The final two layers, which employ 256 filters in the final layer and 128 filters in the third convolutional layer, follow this pattern. A 3x3 kernel size was employed for each layer, and "relu" was selected as activation function because of its capacity to increase the model's non-linearity and efficiency. The pool size for all max pooling layers is 2x2, and the dropout value is 0.2 across all layers. We next add three dense layers after flattening these layers. After the 1st dense layer consisting 256 neurons, a 0.5 dropout layer follows. A second dropout layer with 0.5 follows the 2nd dense layer consisting 128 neurons. Each of the two dense layers used the activation function called "relu." Our final Dense layer comprises 26 neurons that correspond to our 26 classes and makes use of the softmax activation function. We use the Adam optimizer to build our model because it works well for many parameters.

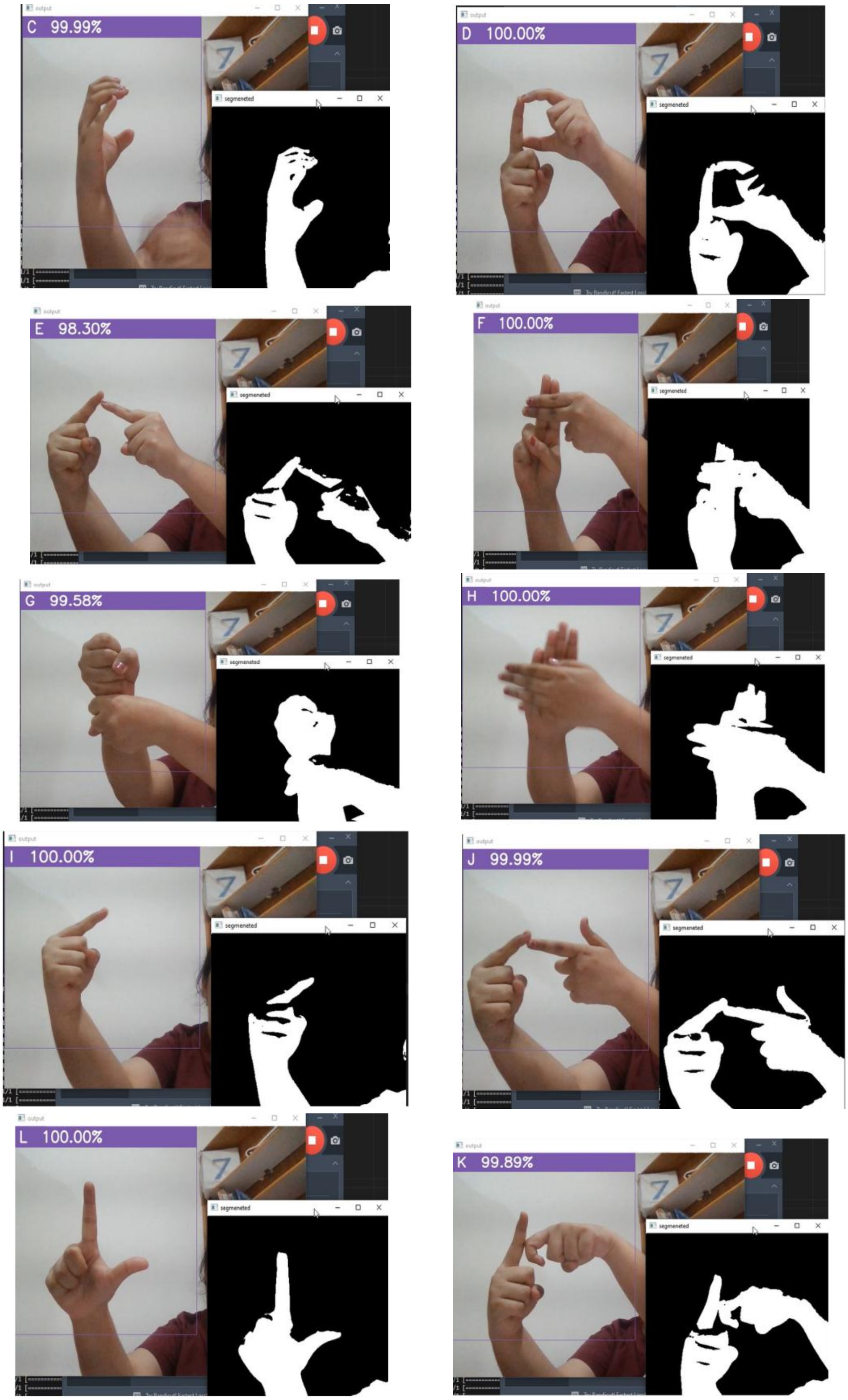
4. LIMITATION AND FUTURE SCOPE

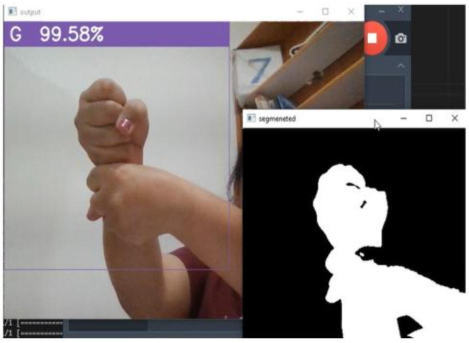
Our existing system can identify 26 letters with a database of 40,000, but due to device constraints, training the model was prolonged and difficult. Consequently, incorporating additional technologies such as GPUs can enhance the process, facilitate expedited training, and result in improved testing and increased opportunities for experimentation. Furthermore, certain environmental limitations, such as background and lighting, occur in our real-time trials; therefore, putting solutions in place to deal with these issues can improve sign language identification systems. Video calls can be made more accessible for those who are deaf or mute by using sign language recognition. A video database of dynamic hand poses can be added to the library of images our model uses to represent static positions. If our model is trained with the right datasets, it can recognize alphabets in various Sign Languages, that involve “American Sign Language (ASL)” and “British Sign Language (BSL)”.

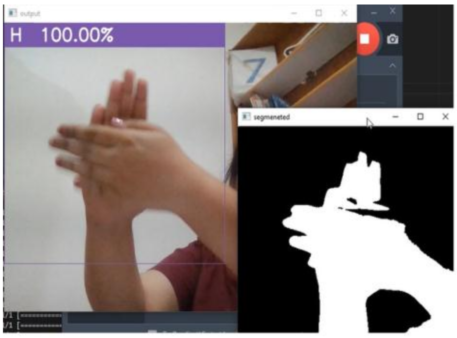
5. RESULTS

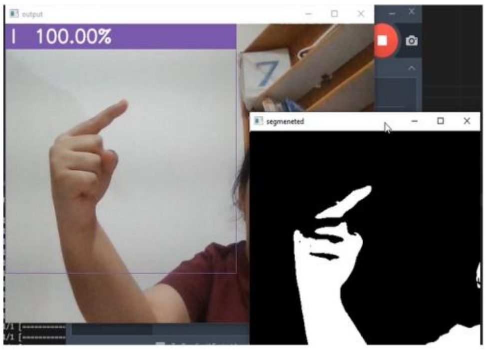


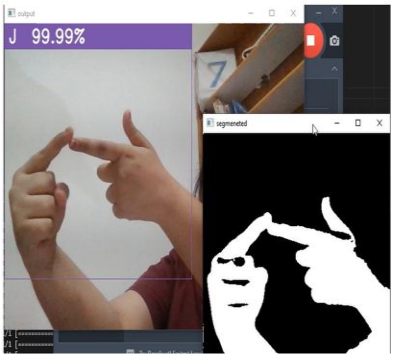


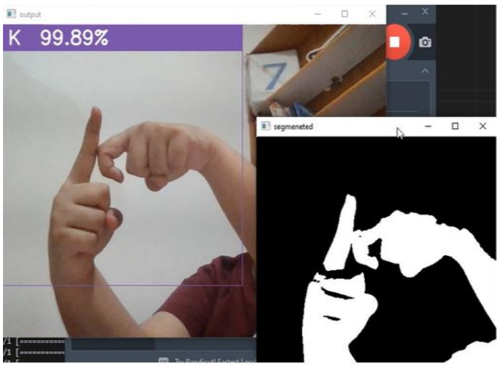

C 99.99%
D 100.00%
E 98.30%
F 100.00%
G 99.58%
H 100.00%
I 100.00%
J 99.99%
L 100.00%
K 99.89%

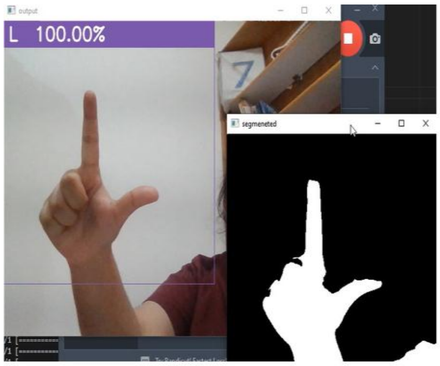

G 99.58%

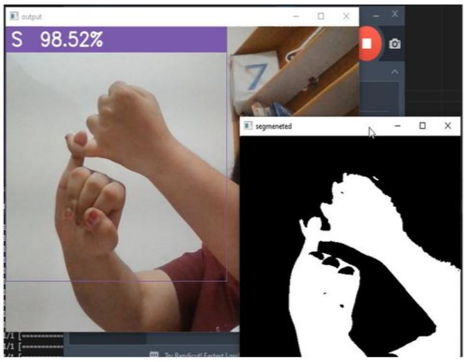

H 100.00%

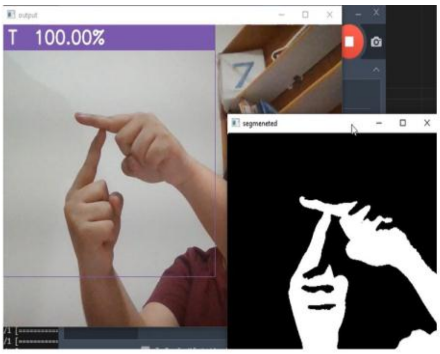

I 100.00%

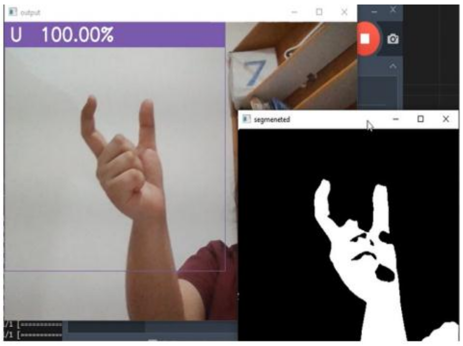

J 99.99%

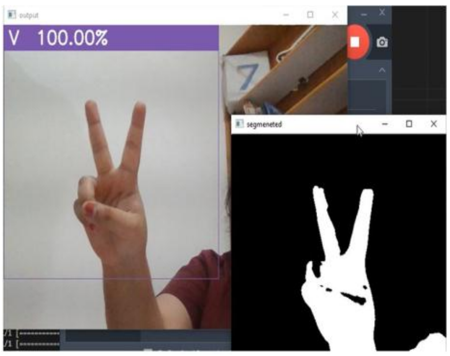

K 99.89%


L 100.00%


S 98.52%


T 100.00%


U 100.00%


V 100.00%

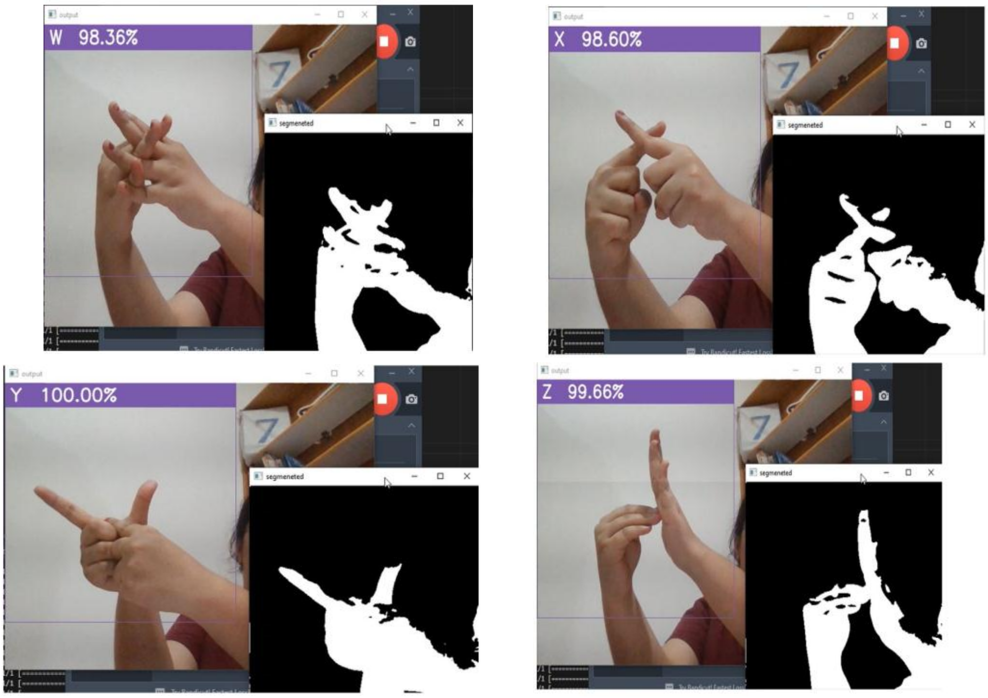


Figure 4: Letters A to Z in ISL with real time accuracy

6. CONCLUSION

In this article, we describe our approach and studies for using a CNN to recognize Indian sign language input in real-time, with an overall model accuracy of 99%. Our model can be scaled to handle larger datasets and provides high-quality and positive identification for both one-handed and two-handed signals done by users with minimum real-time latency. It may be concluded that several layers of Convolutional Neural Networks yield superior accuracy for larger datasets, as evidenced by our custom-created dataset. CNN demonstrates high competence and utility in working with picture datasets, effectively extracting pertinent features for proficient real-time predictions. Through these technical advancements, our research shows that using neural network technologies, like CNN, can help create systems that improve communication for deaf, mute, and disabled groups.

REFERENCES

[1] S.Reshna, “Recognition of static hand gestures of Indian sign Language using CNN,” Published online on 16 April 2020. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

[2] Rajat Soni, “Real-time Recognition Framework for Indian Sign Language Using Fine-Tuned Convolutional Neural Network.” SCRS proceeding of International Conference of Undergraduate Student, 95-106. Published on 2021

[3] Ankita Wadhawan, “Deep learning-based sign language recognition system for static signs.” London: Springer-Verlag, 2020

[4] Harleen Kaur, “Image-based Indian Sign Language Recognition: A practical Review Using Deep Learning Network” unpublished.

[5] Sukanya L, “Indian Sign Language Recognition using Convolution neural Network,” *E3S Web of Conference.391.01058(2023)*

[6] Rachana Patil, “Indian sign Language Recognition using Convolutional Neural Network,”ITM Web of Conferences 40, 03004(2021)

[7] Imran Khan, “Design and Implementation of CNN for Sign Language Recognition.” Jilin Daxue Xuebao (Gong Xue Ban)/Journal of Jilin University (Engineering and Technology Edition), E-publication Online Open Access, Vol:41 Issue:11-22

[8] Chandra Mani Sharma, “Indian Sign Language Recognition Using Fine-Tuned Deep Transfer Learning Model”, unpublished.

[9] Meena Ugale, “A Review on Language Recognition Using CNN”, Published on 2023

[10] Pias Paul, “Modern Approach for Sign Language Interpretation Using Convolutional

Neural Network" published in 2019.

[11] Ilias Papastratis, "Artificial Intelligence Technologies for Sign Language", Published on 30 August 2021.

[12] Pratyush Sinha, "Real Time Sign Language Prediction using CNN and LSTM." volume:05/Issue:04 published in 2023.

[13] Aniket Kumar, "Indian Sign Language Gesture Recognition in Real-Time Using Convolution Neural Network." published in 2021 8th International Conference.

[14] Shashidhar R, "Indian Sign Language to Speech Conversion Using Convolution Neural Network." Published in 2022 IEEE.

[15] Muhummad saad Amin, "A Comparative Review on Applications of Different Sensors for Sign Language Recognition." Published on 2 april 2022.

[16] Dongxu Li, "Word Level deep Sign Language Recognitions from Video: A new Large-Scale Dataset and Method Comparison." Published in 2020.

[17] Sarfaraz Masood, "American Sign Language Character Recognition using Convolution Neural Network." Unpublished.

[18] Sakshi Sharma, "Recognition of Indian Sign Language(ISL) Using Deep Learning Model." Published Online:28 september 2021.

[19] Pranati Rakshit, "Sign Language Detection using Convolutional Neural Network." Published Online:26 march 2024.

[20] Muhammad AL-Qurishi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issue." Published on september 7 2021.

[21] Romala Sri Lakshmi Murali, "Sign Language Recognition System Using Convolution Neural Network and Computer Vision" International Journal of Engineering Innovations in Advanced Technology ISSN: 2582-1431 (Online), Volume-4 Issue-4, December 2022

[22] Abul Abbas Barbhuiya, Ram Kumar Karsh, Rahul jain "CNN based feature extraction and classification for sign language", published: September 2020, Volume 80, pages 3051–3069, (2021)

[23] Shagun Katoch, Varsha Singh, Uma Shanker Tiwary, "Indian Sign Language recognition system using SURF with SVM and CNN", 2022 published by Elservier Inc.

[24] Arpita Halder, Akshit Tayade, "Real-time Vernacular Sign Language Recognition using Mediapipe and Machine Learning International Journal of Research Publication and Reviews Vol (2) Issue (5) (2021)

[25] Eberhard, David M.; Simons, Gary F.; Fennig, Charles D., eds. (2021). Sign language, Ethnologue: Languages of the World (24th ed.), SIL International, retrieved 2021-05-30.

[26] Sharma, A., Sharma, N., Saxena, Y., Singh, A., & Sadhya,D. (2020). Benchmarking deep neural network approaches for Indian Sign Language recognition. Neural Computing and Applications, 1-12.

[27] Yulius Obia , Kent Samuel Claudioa , Vetri Marvel Budimana , Said Achmada, Aditya Kurniawana, "Sign language recognition system for communicating to people with disabilities", (2023)

[28] Katoch, Shagun & Singh, Varsha & Tiwary, Uma Shanker. (2022). Indian Sign Language recognition system using SURF with SVM and CNN. Array. 14. 100141. 10.1016/j.array.2022.100141.

[29] B. Natarajan, 2021, "ISL-CSLTR: Indian Sign Language Dataset for Continuous Sign Language Translation and Recognition" in Mendeley Data (2021)

[30] Pathan, R.K., Biswas, M., Yasmin, S. *et al.* Sign language recognition using the fusion of image and hand landmarks through multi-headed convolutional neural network. *Sci Rep* 13, 16975 (2023). https://doi.org/10.1038/s41598-023-43852-x

[31] I.A. Adeyanjua ,O.O. Bello b , M.A. Adegboye, "Machine learning methods for sign language recognition: A critical review and analysis", Volume 12, November 2021, 200056

[32] Sadia Arooj, Saud Altaf, Shafiq Ahmad, Haitham Mahmoud, Adamali Shah Noor Mohamed, Enhancing sign language recognition using CNN and SIFT: A case study on Pakistan sign language, Journal of King Saud University - Computer and Information Sciences, Volume 36, Issue 2, pa2024,

[33] Ankita Wadhawan, Parteek Kumar, "Deep learning-based sign language recognition system for static signs", Jan 2021.

[34] Necati Cihan Camg, Oscar Koller, Simon Hadfifield and Richard Bowden, "Sign Language Transformers: Joint End-to-end Sign Language Recognition and Translation", 2020.

[35] Abdul Mannan, Ahmed Abbasi, Abdul

Rehman Javed, Anam Ahsan, Thippa Reddy Gadekallu, and Qin Xin "Hypertuned Deep Convolutional Neural Network for Sign Language Recognition", 30 April 2022.

[36] Razieh Rastgoo, Kourosh Kiani, Sergio Escalera, "Word separation in continuous sign language using isolated signs and post-processing", 13 Apr 2022.

[37] Abiyev, Rahib H., Arslan, Murat,"Sign Language Translation Using Deep Convolutional Neural Networks" (2020) *KSII Transactions on Internet and Information Systems*. Korean Society for Internet Information (KSII). doi: 10.3837/tiis.2020.02.009.

[38] D. Bhavana, K. Kishore Kumar, Medasani Bipin Chandra, P.V. Sai Krishna Bhargav, D. Joy Sanjanaa, and G. Mohan Gopi. Hand Sign Recognition using CNN [J]. Int J Performability Eng, 2021, 17(3): 314-321.

[39] Mashael M Asiri, Abdelwahed Motwakel, Suhanda Drar, Robust sign language detection for hearing disabled persons by Improved Coyote Optimization Algorithm with deep learning, AIMS Mathematics, 10.3934/math.2024769, 9, 6, (15911-15927), (2024)

[40] Sumit Kumar, Ruchi Rani, Ulka Chaudhari,"Real-time sign language detection: Empowering the disabled community,"MethodsX,Volume 13, 2024, 102901, ISSN 2215-0161, https://doi.org/10.1016/j.mex.2024.102901