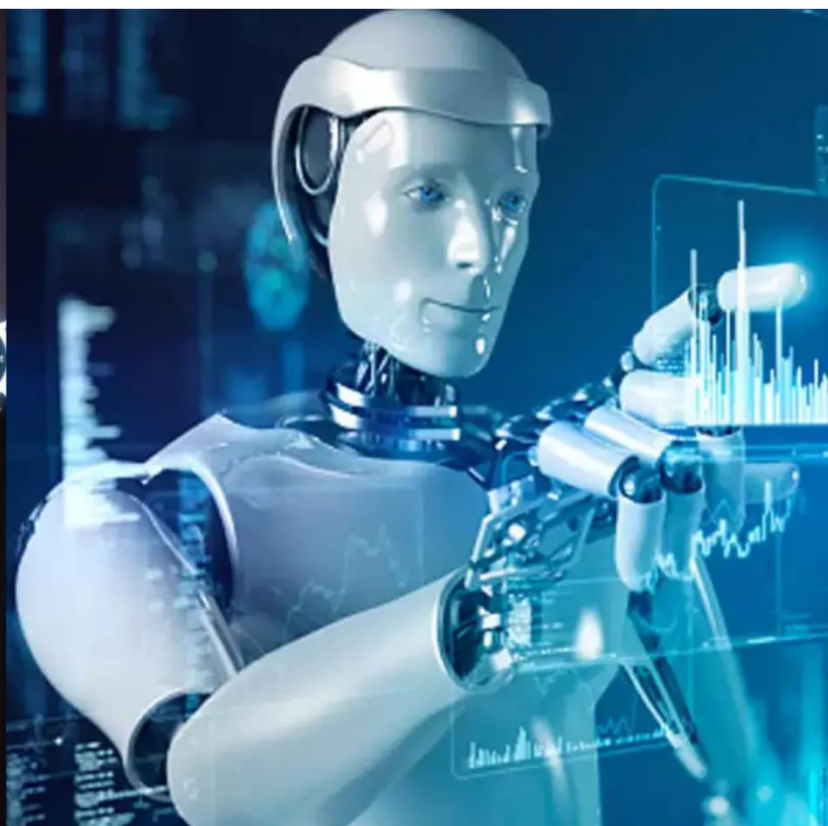




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 8, August 2025

Sign Language Recognition

Arpitha C R, Sunitha G P

P.G. Student, Department of Master of Computer Applications, JNN College of Engineering, Navule, Karnataka, India

Associate Professor, JNN College of Engineering, Navule, Karnataka, India

ABSTRACT: Those of people who depend on sign language for everyday communication are impacted by communication hurdles. This study suggests an instantaneous sign recognition and system that can recognize and interpret the letters A through Z in American Sign Language (ASL) into speech and text. For real-time picture capture and processing, the system makes use of a Feedforward Neural Network built with TensorFlow and Keras, cvzone and OpenCV. The user's hand gesture is recorded by a webcam, pre-processed, and then fed into a CNN structure that have been trained. A text-to-speech engine is utilized to simultaneously translate the output into speech and display it as text. Under controlled illumination conditions, the results of this experiment demonstrate a high accuracy for static hand movements. The model's accuracy on the test dataset is more than 95%. The answer lies in an effective assistive communication tool to the hearing and speech impaired, at least in non-speaking environment.

KEYWORDS: Sign Language, CNN, Computer Vision, Teachable Machine, Mediapipe, Text-to-Speech.

I. INTRODUCTION

Mark language is important because it can enable the hearing and speech defect persons to convey their ideas and also to engage their emotion. However, it is a fact that the general population is not particularly conversant with sign language, thereby there is always a likelihood of encountering barriers to communication in our everyday lives. The predicament creates a necessity to have an automated system which will eventually be in a condition to decode the sign signals into an internationally accepted language consisting of text and speech. Due to the advancement of computer sight and profound understanding, designing intelligent systems that can bridge this divide of communication is now possible. Over the past years a variety of challenges have been developed, aiming to create human computer interaction systems, including sensors, gloves and depth cameras, which appeared high-cost, and could only be applied on a restricted radius. There is, on the other end of the spectrum, the machine learning and vision methods which might offer the cheaper and truly accessible alternative, that is, using such cheap webcams. Gesture recognition: CNNs have also proven rather handy in image classification, and hence can be deployed in gesture recognition. A very powerful and convenient system could be developed to understand the text-to-speech, a collaboration of CNN, and a detection of hand at rest which is used to depict the English alphabet (A-Z) also. This development is primarily aimed at creating and deploying a hand gesture recognition system in real-time on CNN, Tensorflow/Keras, vision frameworks. The system reads and identifies the hand gesture using a web camera, categorises them and reference the predicted letter to the screen plus a text to speech system pronounces the predicted letter to sound. This solution is the quick, effective and inexpensive aiding talent that is not only comprising of inclusive communication, but it establishes a foundation of subsequent modifications such as: constant sign identification, word prediction, multi lingual support and so on.

II. LITERATURE SURVEY

[1]. Rupesh Kumar, Ashutosh Bajpai, and Ayush Sinha (2023) constructed a real-time ASL alphabet recognition system based on the Mediapipe as a feature extractor and a CNN classifier. Their algorithm was 99.95% accurate which indicates good preprocessing and performance. [2] In a review of the approach to vision-based hand gestures, M. Oudah (2020) has compared the technique to the RGB technique, depth techniques, and sensor-based techniques. Their weaknesses and strengths were described, which gave context to webcam-only CNN systems. [3] Cheng et al. (2020) proposed an FCN that has a GFE module to acquire spatial and temporal representations of sign language. They work with real-time continuous recognition without pre-segmentation. [4] Li and Meng (2022) introduced a Transformer based multi-view spatial-temporal network using CTC decoder. Their model demonstrated high feature synergy with very short term mistake rates in continuous sign language recognition.

[5] V. Adithya et al. (MediaPipe hand-tracking 2020) Herein, we have presented deep grip, Deep CNN, a deep hand gesture recognition, which enhanced the performance but increased depth of the network, however, the capacity to increase accuracy with scale and variability depended on the dimensions of the data set. [6] Aziza Zhidebayeva et al. (2024) embedded two-stream context-aware 2D CNN-based real-time fingerspelling recognition. The technique enhanced the intensity of different lighting and background situations. [7] Arnav Gangal et al. (2023) trained ASL character detectors on dual-input CNNs and reported that tuning, augmentation, and hand landmarks could improve accuracy to 96.42. [8] Elboushaki et al. (2020) presented MultiD-CNN, a model that is trained to learn features using RGB-D sequence of gestures. It utilizes both the altitudinal and depth information to increase the level of robustness over RGB-only systems. [9] Aditya Raj Verma et al. (2024) This study combined with CNNs to create an efficient, real-time ASL detection system using webcam feeds, achieving an accuracy of 99.12%. [10] Bader Alsharif et al. (2023) tested five ASL recognition models, the most accurate of which is ResNet-50 with 99.98 percent accuracy.

III. METHODOLOGY

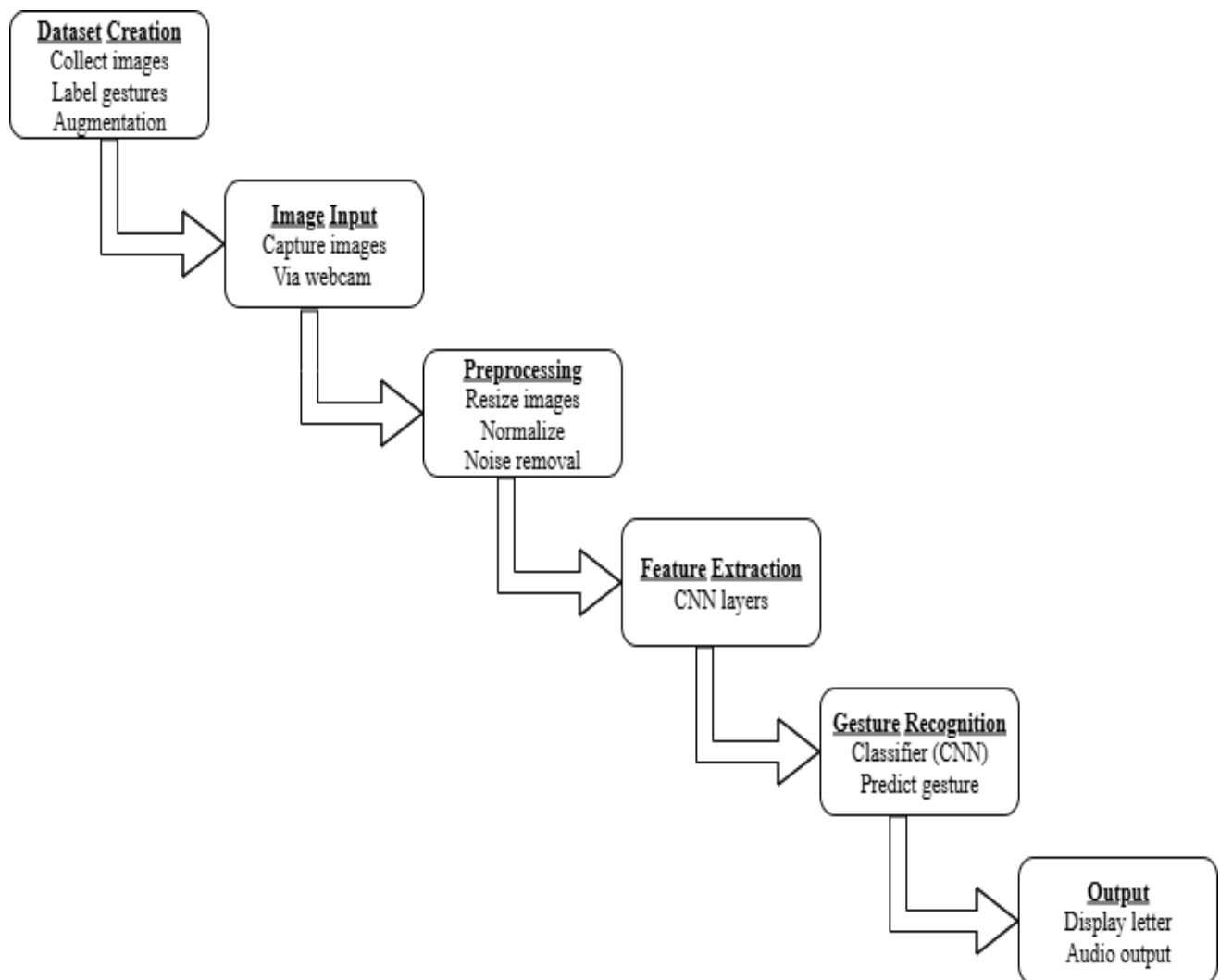


Fig 1: System Flow

Dataset creation: In demand to assist the prototypical train and detect gestures effectively, dataset creation entails gathering hand gesture photos, classifying them, and adding augmentation such as rotation or flipping to boost diversity. Dataset = Collect Images + Label Gestures + Augmentation

Image Input: Image preprocessing is the procedure when unprocessed images are transformed for better analysis by the model. Resizing ensures all photos have the same dimensions, making them consistent for input. Normalization speeds up and stabilizes learning by scaling pixel values, typically between 0 and 1. By removing undesired distortions or aberrations, noise reduction enhances image clarity and aids the model in concentrating on key elements.

Image preprocessing: By shrinking photos to a consistent size and normalizing pixel values ($I_{norm} = I/255$) for quicker training and employing filters to eliminate noise improve clarity and concentrates on the hand gesture.

Feature Extraction: CNN feature extraction automatically finds significant patterns, such as forms and edges. The CNN layers-convolution, activation (ReLU), and pooling- allow the model to learn intricate pattern for precise gesture detection by detecting features, adding non-linearity, and reducing dimensions.

Gesture Recognition: CNN-extracted features are leveraged in gesture recognition, each indication is given a probability and the CNN classifier predicts which gesture has the highest probability.

Output: The user is shown the recognized gesture at the output stage. The technology makes the interaction understandable and accessible, particularly for communication reasons, by displaying the matching letter on the screen and producing an audio output by turning the anticipated letter into voice.

IV. EXPERIMENTAL RESULTS

Figures show the results of translating Hand Gestures to Text. Fig 1 shows the detected letter Z. fig 2 shows the detected letter C. fig 3 shows the detected letter Q. fig 4 shows the detected letter O. fig 5 shows the detected letter B.

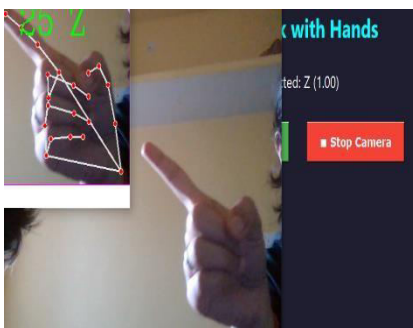


Fig 2a: Detected Letter Z

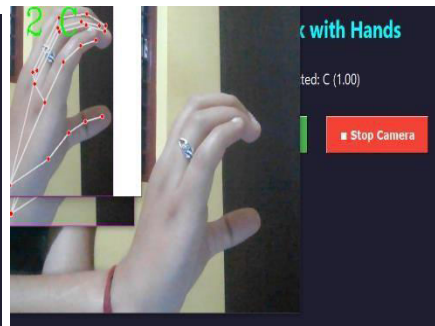


Fig 2b: Detected Letter C

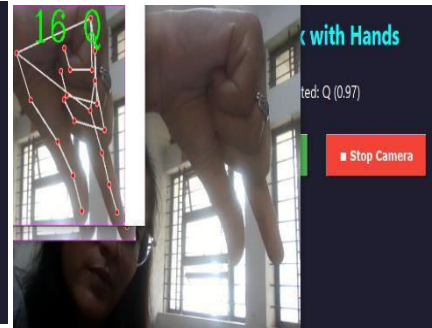


Fig 2c: Detected Letter Q

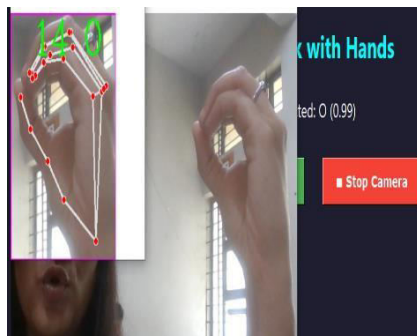


Fig 2d: Detected Letter O

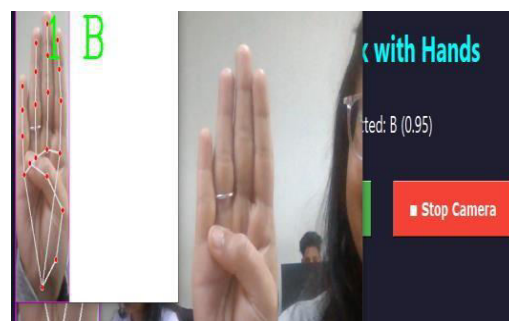


Fig 2e: Detected Letter B

Figure 2: Detecting the Letters

These images look to be screenshots from a computer vision or hand-tracking application. Each one illustrates a result where the system has identified hands and is attempting to identify a certain gesture, referred to as a certain gesture, referred to as a “tech.” Each detection’s confidence level is indicated in parenthesis, with 1.00 representing the highest level of confidence.

The majority of screenshots display the command to “Stop Camara,” which was probably what the detected gesture caused. For instance, one image fully confidently identified a “Z” gesture, while others identified a “C”, “Q” and one in which the system was 99% certain a gesture was done but the precise letter was not identified. It was able to identify “B” at the top before the hand was spotted. To put it simply, the system is keeping an eye out for hand gestures and upon detected them stopping the camara.

Table 1: Letters and their Accuracy

Letters	Accuracy
B	0.95
C	1.00
O	0.99
Q	0.97
Z	1.00

V. CONCLUSION

In this research, we developed a system for hand signal identification that can convert the A-Z gesture alphabet into speech and text. Hand signs are captured by a camara, processed by machine learning models, and the predicted letter is displayed on the screen along with audio output. Through self-expression without the use of an interpreter, this project facilitates for those who are profoundly deaf in interaction. Despite many drawbacks, such as problems with lighting and background, it is a helpful first step in developing more advanced sign language translation systems and enhancing accessibility for all.

REFERENCES

1. Kumar, R., Bajpai, A., & Sinha, A. (2023). Mediapipe + CNN real-time ASL recognition.
2. M. Oudah, “A comprehensive literature review on vision-based hand gesture recognition methods: RGB, depth, and sensor-based paradigms.” 2020.
3. X. Cheng, “Continuous recognition of sign language, end to end, based on a fully convolutional network and optimization of gloss features.” 2020.
4. R. Li and L. Meng, “Multi-View Spatial-Temporal Network for Continuous Sign Language Recognition,” 2022.
5. V. Adithya, et al., “Deeo Gesture: A deep CNN architecture for static hand gesture recognition,” 2020.
6. A. Zhidebayeva, G. Nurmukhanbetova, S. Aldeshov, K. Zhamalova, S. Mamikov, and N. Torebay, “Context-aware two-stream 2D deep CNN for real-time fingerspelling recognition.
7. A. Gangal, A. Kuppahally, and M. Ravindran, “Ablation study on ASL character detection using dual-input CNNs,” 2023.
8. Elboushaki, A., et al. (2020). MultiD-CNN: Feature learning from RGB-D gesture sequences.
9. A. R. Verma, “Real-time American Sign Language detection using MediaPipe.” 2024.
10. A. Elboushaki, “MultiD-CNN: Feature learning from RGB-D gesture sequence.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details