



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 11    **Issue:** IX    **Month of publication:** September 2023

**DOI:** <https://doi.org/10.22214/ijraset.2023.55871>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# American Sign Language Recognition and its Conversion from Text to Speech

Aditi Bailur<sup>1</sup>, Yesha Limbachia<sup>2</sup>, Moksha Shah<sup>3</sup>, Harshil Shah<sup>4</sup>, Prof. Atul Kachare<sup>5</sup>

Computer Engineering Department, Shah & Anchor Kutchhi Engineering College

**Abstract:** *Speech impairment is a complicated condition that impairs a person's capacity for verbal and audible communication. Those who are impacted frequently use sign language and other alternate forms of communication. While sign language has gained popularity, bridging the communication gap between those who sign and those who don't remains a challenge. Our project endeavors to address this issue by developing an innovative application that offers real-time sign language-to-text translation. This technology aims to facilitate seamless communication between those who use sign language and those who do not. To achieve this, we have constructed a cutting-edge sign language recognition system as part of our project. Our system primarily utilizes American Sign Language (ASL) as its foundation. To accurately detect gestures, we employ a Convolutional Neural Network (CNN) with Inception V3 as the underlying model. This project's core objective is to harness Machine Learning Techniques for converting ASL hand gestures into text and vice versa. We go beyond mere translation by enabling real-time American Sign Language interpretation through single-hand gestures. Furthermore, our system can recognize ASL words, converting them into text before rendering them into audible speech.*

**Keywords:** *Sign language detection, American sign language, text-to-speech translation, CNN, Inception V3.*

## I. INTRODUCTION

People with hearing or speech impairments often rely on sign language, a visual means of communication, to interact with each other and the broader community. 400 million individuals worldwide have hearing impairment, according to the World Health Organization. Recent advancements in research aim to enhance communication accessibility for individuals with disabilities. For those who are deaf or mute, sign language recognition systems act as invaluable interpreters. These systems play a crucial role in converting sign language into understandable text. Leveraging imaging technologies, these systems work diligently to identify sign gestures and translate them into text comprehensible to the deaf and mute community. However, there are inherent challenges in this endeavor, primarily stemming from the diversity of spoken languages across regions and nations. American Sign Language (ASL), for example, employs 22 distinct forms to represent the 26 letters of the alphabet and single-handed signs for numbers. ASL, like spoken languages, is a complete and natural language, sharing numerous linguistic traits.

This paper's primary objective is to enhance the recognition of ASL sign gestures using an advanced neural network model, improving upon prior research. We employ a Convolutional Neural Network (CNN) to develop ASL hand gesture recognition software, which outperforms many existing models, offering greater accuracy in sign language interpretation for the deaf and mute community.

### A. Contribution of the Paper

- 1) We list the important ways to turn sign language into text.
- 2) We use computer programs to make different models.
- 3) We evaluated the models and chose the best one.
- 4) Show the complete working of the best model.
- 5) Further convert the generated text to speech.

## II. RELATED WORK

In the context of recognizing American Sign Language (ASL), several contemporary methods and technologies have been explored: Paper [1] uses neural networks to help people who are deaf communicate with those who don't understand sign language, focusing on American and Indian Sign Language. It trains a three-layer neural network to recognize sign language and translate it into English. [2] This "Sign Language Translator" uses neural networks to convert spoken language into American Sign Language, making communication easier for Deaf-Mute individuals and aiding in language teaching. [3]

Using a trained neural network, this paper achieves 90% accuracy in recognizing ASL in real-time. It uses a unique approach involving skin tone calculation for hand gestures.[4] An application is developed that translates ASL into text and then into speech in real-time via a computer's webcam. Convolutional neural networks are used for gesture detection with high accuracy. [5] This research employs convolutional neural networks to recognize static American Sign Language images. It's trained and validated on a dataset containing images of English alphabet signs. [6] This paper introduces a system that recognizes hand signs and generates Bangla speech with a CNN-based model, achieving 92% accuracy.[7] This work focuses on recognizing objects in images and generating speech based on them. It uses various techniques like SIFT, SURF, and HOG for feature extraction, SVM for object recognition, and HMMs for speech generation.



Fig. 1: ASL Dataset

### III. DATASET

The foundation of our network's training lies in the ASL Alphabet collection, which encompasses a diverse set of ASL signs. Within this collection, we have amassed a total of 87,000 images, each standardized at 200x200 pixels. These images are categorized into 29 distinct classes, encompassing the 26 English alphabets, alongside three supplementary signs denoting space, delete, and nothing. To fortify our model's ability to handle real-world scenarios, we adopted data augmentation techniques. These improvements included brightness adjustments that introduced fluctuations of up to 20% in low-light situations and zoom shifts that allowed photos to be zoomed out by up to 120%. These augmentations enable our model to excel in a broader range of environmental conditions. For the purpose of model validation, we carefully selected 28 images from this extensive collection. These validation images were set aside to assess the model's performance. The remaining images were then utilized for the rigorous training of our ASL alphabet recognition model.

Figure 1 provides a visual glimpse into our dataset, showcasing a selection of sample images that represent the rich diversity of ASL signs our model has been trained to recognize.

### IV. PROPOSED MODEL

In our pursuit of developing an effective solution, we delved into transfer learning to glean insights into the task at hand. However, it's noteworthy that our network was ultimately crafted from the ground up. The cornerstone of our engineering endeavor revolves around a Convolutional Neural Network (CNN) architecture. This architecture is characterized by a multi-convolutional design that incorporates densely connected layers. The design consists of two sets of completely attached layers, separated by a dropout layer, leading to a final output layer. Additionally, there are four pairs of convolutional layers, all the layers are tailed by dropout and max-pooling layers.

This meticulously designed model forms the bedrock of our approach, poised to address the challenges at hand with precision and efficacy.

## V. METHODOLOGY

Employing Transfer Learning and Data Augmentation, we harness powerful techniques to construct a deep learning model for the American Sign Language dataset.

### A. Transfer Learning

This technique, known as transfer learning, leverages a model originally designed for one task as the foundation for another. In the realm of deep learning, it proves highly valuable, allowing us to utilize pre-trained models, thus conserving substantial computational resources and time. Its remarkable performance benefits make it particularly advantageous for complex challenges in Artificial Intelligence and NLP (Natural Language Processing).

### B. Model Architecture

Our neural network is built upon Google's Inception v3 model. To fine-tune this model, we lock the initial 248 layers i.e. till the 3rd last block, and exclusively train the model using the last two blocks. Furthermore, we replace the completely attached layers at the peak of the Inception network with a new customized set of attached layers. Our unique architectural design includes two completely attached layers: one featuring 1024 Rectified Linear units, and the other with 29 Smooth arg max units, tailored for foreseeing 29 distinct ASL sign classes. We then train the model using a fresh batch of ASL images specifically curated for our application.

### C. Application Integration

After successfully training the model, it is seamlessly integrated into the application. We utilize OpenCV for frame extraction from a video stream. Within the system's interface, a defined portion of the screen is marked by a colored rectangle where in we display signs for detection and identification. The model meticulously analyzes the captured frames, making sign predictions based on the displayed hand gestures. Predictions are classified into three confidence levels: for signs with low confidence (between twenty to fifty percent confidence), display a "Maybe [sign] - [confidence]" result on the screen, and high certainty signs (over fifty percent confidence) are presented as "[sign] - [confidence]" outputs. In this scenario, [sign] stands for the sign predicted by the model, and [confidence] measures the level of certainty the model has for that specific prediction. When confidence falls below 20%, the model refrains from producing any output.

### D. Speech Conversion

Upon successfully recognizing the signs, the identified text is forwarded to the Google Speech Conversion API. This API transforms the text into audible speech, enhancing accessibility for individuals with hearing impairments. Through the implementation of this comprehensive methodology, we enable efficient ASL sign recognition and seamless communication, effectively integrating machine learning and computer vision techniques.

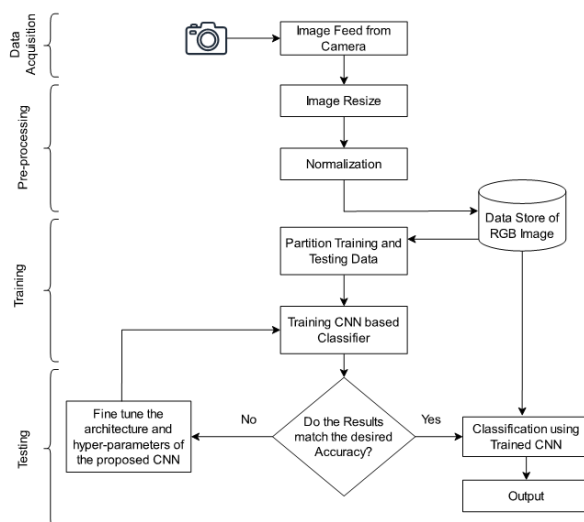


Fig. 2: Flowchart of the proposed system



### VI. EXPERIMENT

In our experimental endeavor, we've harnessed the synergy of TensorFlow's backend seamlessly integrated with Keras. Dedicated functions and models for neural networks and image processing are available through Keras. Its design is centered on facilitating swift experimentation, featuring user-friendly functions for crafting custom neural network models, as well as the flexibility to implement and fine-tune pre-existing ones. The model we employed underwent rigorous development and testing within the confines of Google Collaboratory—an invaluable research tool generously provided by Google for machine learning exploration. This environment prioritizes GPU support, significantly curtailing training durations and enhancing model development efficiency. To bolster the robustness of our Convolutional Neural Network (CNN) model, we judiciously applied data augmentation techniques during training and validation phases. The outcome of our experimentation is vividly illustrated in the precision and loss graphs, depicted in Figure 3 and Figure 4, respectively. These graphs unmistakably reveal our model's consistent and commendable progress, with discernible enhancements in precision and notable reductions in loss, particularly when considering augmented data.

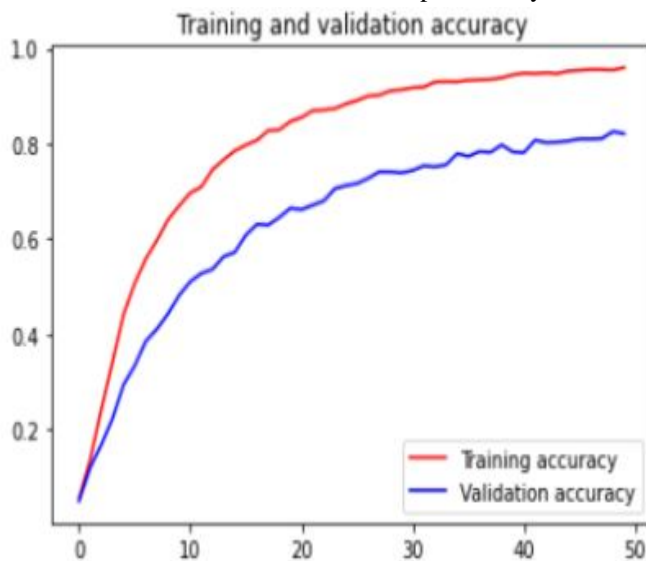


Fig. 3: Compares accuracies of training and validation

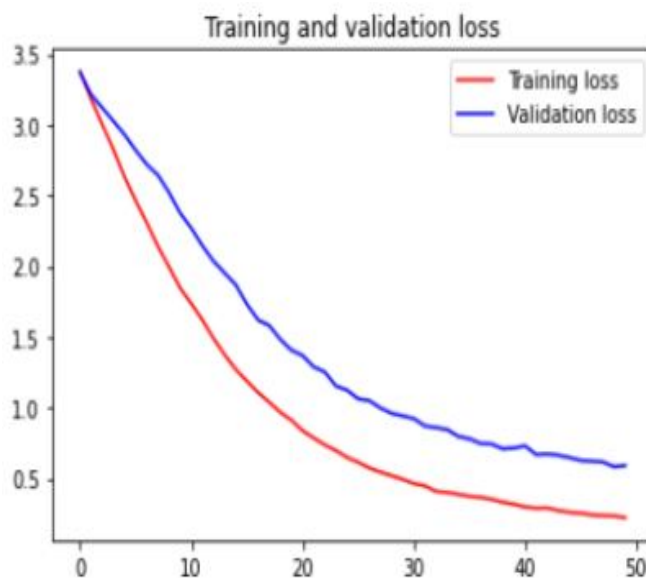


Fig. 4: Compares loss of training and validation

### VII. RESULT AND ANALYSIS

TABLE I  
ANALYSIS OF SYSTEMS PERFORMANCE

No.	Alphabets	Test1	Test2	Test3	Test4	Performance (%)
1	A	✓	✓	✓	✓	100
2	B	✓	✓	✓	✓	100
3	C	✓	✓	✓	✓	100
4	D	✓	✓	✓	✓	100
5	E	✗	✓	✓	✓	75
6	F	✓	✓	✓	✓	100
7	G	✓	✓	✓	✓	100
8	H	✓	✓	✓	✓	100
9	I	✓	✓	✓	✓	100
10	J	✓	✓	✓	✓	100
11	K	✓	✓	✓	✓	100
12	L	✓	✓	✓	✓	100
13	M	✓	✓	✓	✓	100
14	N	✗	✓	✓	✓	75

After creating and analyzing the model, we found the training accuracy and loss to be 0.98 and 0.12 respectively, and validation accuracy and loss to be 0.949 and 0.19 respectively. The overall test accuracy was 96%.

The conclusive outcomes presented above unequivocally demonstrate the impressive performance of our model. Notably, our model attains a remarkable validation accuracy, surpassing many of its counterparts. Specifically, our model achieves an outstanding validation accuracy of 95%, a notable improvement when compared to the majority of existing models.

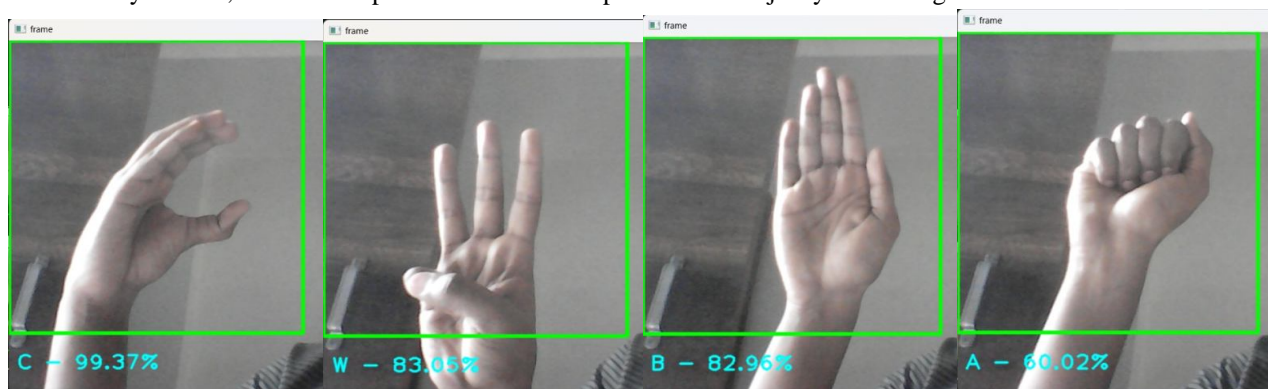


Fig. 5: Testing Set Results.

### VIII. CONCLUSIONS

This study showcases an American Sign Language (ASL) classification algorithm built upon the foundations of deep learning. Our novel approach offers an efficient solution and leverages the simplicity of a basic camera as a readily available dataset source. Crucially, our system thrives on a continuous influx of meaningful training data, seamlessly integrated into the processing pipeline outlined above. This adaptability not only ensures robustness but also contributes to a scalable solution, catering to the growing need for accessible camera technologies.

Remarkably, our model architecture surpasses its predecessors in both training precision and validation accuracy. Moreover, the proposed design exhibits a lower overall training and validation loss, underscoring its efficiency. The pinnacle of our achievement lies in the remarkable recognition rate of our suggested model, soaring to an impressive 96.43%. This performance exceeds that of cutting-edge classifiers, solidifying the significance and effectiveness of our approach in ASL classification.

### REFERENCES

- [1] Murat Taskiran, Mehmet Killioglu and Nihan Kahraman, "A Real-Time System for Recognition of American Sign Language by using Deep Learning", 2018, 41st International Conference on Telecommunications and Signal Processing (TSP), DOI: 10.1109/TSP.2018.8441304.
- [2] Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur, Dr. Dayananda P, "Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network," INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY, 2020, DOI : 10.17577/IJERTCONV8IS15042
- [3] Aakriti Rustagi, Shaina and Neha Singh, "American and Indian Sign Language Translation using Convolutional Neural Networks," IEEE, 2021, 8th International Conference on Signal Processing and Integrated Networks (SPIN) , DOI: 10.1109/SPIN52536.2021.9566105
- [4] Aakriti Rustagi, Shaina and Neha Singh, "American and Indian Sign Language Translation using Convolutional Neural Networks," IEEE, 2021, 8th International Conference on Signal Processing and Integrated Networks (SPIN) , DOI: 10.1109/SPIN52536.2021.9566105
- [5] Ahmed, M. Islam, J. Hassan, M. U. Ahmed, B. J. Ferdosi, S. Saha, M. Shopon et al., "Hand sign to bangla speech: A deep learning in vision based system for recognizing hand sign digits and generating bangla speech," IEEE, 2019
- [6] Ashvini Butte, Sarita Jadhav, Sayali Meher,"Image feature extraction, classification, recognition done using MATLAB and conversion to speech using HMM", IEEE, 2020.
- [7] "Sign language recognition based on HMM/ANN/DP",International journal of Pattern Recognition and Artificial Intelligence,DOI:10.1142/S0218001400000386
- [8] Rajaganapathy, S. and Aravind, B. and Keerthana, B. and Sivagami."Conversation of Sign Language to Speech with Human Gestures",Procedia Computer Science, 2015.
- [9] Kausar, Sumaira and Javed, Muhammad and Tehsin, Samabia and Anjum, Muhammad Almas, "A Novel Mathematical Modeling and Parameterization for Sign Language Classification", International Journal of Pattern Recognition, 2016.
- [10] P. Vijayalakshmi and M. Aarthi, "Sign language to speech conversion",International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, 2016.
- [11] Warriar, Keerthi and Sahu, Jyateen and Halder, Himadri and Koradiya, Rajkumar and Raj, V, "Software based sign language converter",ICCSPP, 2019.
- [12] Hasan, Mokhtar and Mishra, Pramod, "HSV Brightness Factor Matching for Gesture Recognition System",International Journal of Image Processing, 2010.
- [13] Itkarkar, R. and Nandi, Anil, "Hand gesture to speech conversion using Matlab",International Conference on Computing, Communications and Networking Technologies, ICCCNT, 2013.





10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)