

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 3, June 2025



# ML-Based Indian Sign Language Translator for Speech-Impaired Individuals

Prathmesh Patil<sup>\*1</sup>, Shreya Pansare<sup>\*2</sup>, Mayur Bhagade<sup>\*3</sup>, Ms. P. B. Gholap<sup>\*4</sup>

<sup>\*1</sup>Dept. AI & DS Engg, Jaihind College of Engineering Kuran, Pune

<sup>\*2</sup>Dept. AI & DS Engg, Jaihind College of Engineering Kuran, Pune <sup>\*3</sup>Dept. AI & DS Engg, Jaihind College of Engineering Kuran, Pune

<sup>\*4</sup>Dept. AI & DS Engg, Jaihind College of Engineering Kuran, Pune

**Abstract**: The people with speech disabilities usually face problems interacting with normal people, and there is a need to find a way to make communication with normal human beings easier. This paper introduces a system that recognizes static and real time signs gestures and translates them into text and speech. The system uses image processing and machine learning to identify hand signs from a dataset. After recognizing the gesture, it converts it into text and then into speech using text-to-speech technology. The system was tested on a standard dataset and showed an accuracy of 99% making it effective and reliable. This tool aims to make communication easier and more accessible for everyone. This system can be improved to recognize dynamic gestures and support more languages

Keywords: Sign language, Convolutional neural Network, LSTM, Landmark.

### I. INTRODUCTION

Communication is an important aspect of human interaction, enabling individuals to express their thoughts, emotions, and needs. However, for speech-disabled individuals, expressing themselves can be difficult, often leading to social and professional barriers[1]. Indian Sign Language (ISL) serves as a primary mode of communication for the speech-disabled individuals in India. Sign language is not just a gesture using fingers and palms; it involves visual cues through the eyes, face, mouth, eyebrows, etc. Additional components, like facial expressions, involve expressing the complex meaning [2] [6].Despite its importance, the limited understanding and acceptance of ISL among the general population create a significant communication gap. A Sign Language Recognition and Translation System using Machine Learning (ML) aims to bridge this gap by converting sign language gestures into text or speech, enabling real-time communication between speech-disabled individuals and others. By leveraging computer vision, deep learning, and natural language processing (NLP), this system can recognize hand gestures, facial expressions, and body movements to interpret ISL accurately[6].

In this process we are trying to build a model that must be able to guess the sign from the frame that a webcam is capturing and classify it into the classes it is trained under. The translation must be smooth and must not be time consuming. We tried to reduce the time complexity and must be able to classify similar kinds of signs with efficiency. To do so the two models are employed one is with image model and the other one is the landmark model which is based on the media pipe's hand landmark detection mod

### II. DATASET

There are multiple datasets used in the training phase of the model. There are special datasets on Indian Sign Language (ISL) and some of them are gathered from Kaggle which were used in previous papers which were used for reference purposes. Those images were already in the 128 \* 128 resolutions. There is another dataset which was created by ourself for the training process. The images were 1280 \* 720 resolution. These images were directly used to train the landmark model for better detection of the hand landmarks with better quality and then those images were converted to 128 \* 128 resolution for training the image model which reduces the time complexity for the model. The images were even processed with different augmentation and the different processing. The processing included adding more

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/568





International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

#### Volume 5, Issue 3, June 2025



brightness and adding different kinds of rotations in the image. This could help in making the model more robust even if used in different situations. The images were captured with different backgrounds to make the model more reliable. There is image (3.1), which is a representation of what kind of dataset used to train the models.



Figure 1: The sample from Datset

### **III. METHODOLOGY**

The methodology for the project is to divide the recognition in the two parts. The first one is the Convolutional Neural Network (CNN) image model and the second one is Hand Landmark model which is used to classify the landmarks into the classes of alphabets in the sign language.

#### A. Image Model

Firstly, there is a dataset which is used to train the model for predictions. To train the model we are processing the dataset in such a way. The preprocessing stage takes the input as an image and convert the image in the 128 \* 128 size to make an image smaller and in a normal size for processing. After this the image is normalized by dividing by 255 to scale the image pixels from 0 to 255 for better preprocessing. The model has 3 different blocks for feature extraction, global average pooling and fully connected classification layers. Each layer contains all the conv2D layers for feature extraction. Every layer in the block has a Relu activation and Batch regularization to train the model efficiently. The model faced some overfitting as some of the signs are similar to each other. Such as the sign for letter D and P are somewhat similar so it occurred overfitting and was not able to train as efficiently. To overcome this problem, we used L2 regularization. This approach reduced the possibility of overfitting. The kernel size set to 3\*3 with the same padding to preserve the spatial dimensions.

The first block has the 32 filters and two convolutional layers and a MaxPooling layer which removes all non-important features and concentrates on only with higher values for further processing. To improve the generalization of the model the dropout layer is used which drops the co-adaptations of the neuron to reduce the overfitting. The use of batch normalization stabilizes the training process and makes the learning quicker without losing features. There is one more layer called the Global Average Pooling layer which is used to reduce the number of parameters and mitigate the overfitting concerns for the model. There are fully connected classification layers total in 128. The dropout of 60% is introduced.

While training the model the stable learning rate of 0.0001 is used for stable training. This approach is useful for better training and fast processes. Not making the preprocessing much complex and making it look normal while only processing the image with the sign is making the preprocessing faster and optimal. The image model is trained on the vast dataset which can make the model make predictions with much confidence. This makes the processing even more reliable.

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/568





International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 3, June 2025



#### B. Landmark Model

The second part is the landmark model which plays one of the important roles in prediction of the sign. The structure of the model is the same compared with the image classification model. The image preprocessing is quite different from the image model as it has nothing to do with the image resolution but hands. Used Mediapipe's hand landmark detection model for the detection of the landmarks. The Mediapipe's hand landmark detection model returns 21 landmarks for one hand detected in the frame and for two hands it returns 63 landmarks. The model is in such a way that it extracts landmarks from the frame if only one hand is available for the as it is for some signs only one hand is available so only one hand landmarks are appended and for the other hand all zeros are appended. Which makes the learning even efficient. The model also has a global average pooling layer to make the learning of the model even more stable and reliable. The model also contains the same kind of structure as the image CNN model and also contains dropout layers to avoid overfitting. The learning rate is kept to the 0.0001 so it has a stable weight assigned for the classes. It too has 128 fully connected classification layers which are connected to 26 classification layers.



Figure 2: Architecture of image model

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/568





International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 3, June 2025



### C. Meta Model

The meta model is a model kind of structure based on the image model and landmark model. It takes inputs as a final prediction from both the models and processes those two to reach the final decision of the class. The image model is a superior model and the landmark model is treated as an associate model to help the model to reach the conclusion. The confidence for the image model must be either 0.5 or above and the confidence for the landmark model must be 0.6 or above and also if both the models have the same prediction then the prediction is termed as right and final prediction otherwise ambiguity occurs and the result is not displayed or stored.



Figure 3: Architecture of models and result finalization

### **IV. RESULTS**

The models are trained on the various datasets, and the results are different for both the models. The image model trained on the images and the images are converted to 128 \* 128 resolution. The model is based on convolutional neural networks (cnn) algorithms with conv2D layers in three different blocks. The model has around 0.9978 accuracy and the value accuracy of 0.9998. The information loss is around 0.2 which is low. The model was overfitting in some ways with ambiguity in understanding some of the signs which are similar to each other. The model performs well on the unseen data after training on multiple datasets with different distributions. The second model is the model of landmarks. The landmark model is trained on the same dataset which was used to train the language model. The model was able to understand the variability in the landmarks for each sign. The landmarks extraction was based on the mediapipe's hand landmark detection model. It returns 21 landmarks for a single hand and 63 for two hands.

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/568





International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

#### Volume 5, Issue 3, June 2025



The model training accuracy was 97% and the value loss is around 0.4 but when treated with the unseen data it gets accuracy around 96.6% and with the same amount of the value loss. The model has the better learning ability and is able to get the difference between the two similar signs comparatively better.

The meta model is a combination and it only gathers the predictions of other two models and checks if the predictions are reliable or not. This step increases the accuracy and makes the model even better performing.

### V. CONCLUSION

In the conclusion we can state that the approach used with image feature extraction and landmark extraction has a good amount of accuracy with optimal time complexity. The approach suggests to focus on the more off feature extraction but using landmarks for the same can be even more beneficial for time complexity. For better accuracy the approach can be more based on the landmarks so for the live or continuous sign detection it might perform better. The conclusion is that using a combination of the both the models simultaneously can be more beneficial. For future betterment one can train a model on vast data then the accuracy can even better.

#### REFERENCES

[1] S. M. Miah, M. A. M. Hasan, S. Nishimura and J. Shin, "Sign Language Recognition Using Graph and General Deep Neural Network Based on Large Scale Dataset". in IEEE Access, vol. 12, pp. 34553-34569, 2024, doi:10.1109/ACCESS.2024.3372425.

[2] Desai, A., Berger, L., Minakov, F., Milano, N., Singh, C., Pumphrey, K., Ladner, R., Daum'e III, H., Lu, A.X., Caselli, N. and Bragg, D."ASL citizen: a community-sourced dataset for advancing isolated sign language recognition". 2024

[3] R. Zuo, F. Wei and B. Mak, "Natural Language-Assisted Sign Language Recognition," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 14890- 14900, doi: 10.1109/CVPR52729.2023.01430.

[4] Zhou, Benjia, Zhigang Chen, Albert Clap'es, Jun Wan, Yanyan Liang, Sergio Escalera, Zhen Lei, and Du Zhang. "Gloss-free sign language translation: Improving from visual-language pretraining." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.20871-20881. 2023.

[5] R. Kothadiya, C. M. Bhatt, T. Saba, A. Rehman and S. A.Bahaj, "SIGNFORMER: DeepVision Transformer for Sign Language Recognition," in IEEE Access, vol. 11, pp. 4730-4739, 2023, doi:10.1109/ACCESS.2022.3231130.

[6] D. R. Kothadiya, C. M. Bhatt, A. Rehman, F. S. Alamri and T.Saba, "SignExplainer: An Explainable AI-Enabled Framework for Sign Language Recognition With Ensemble Learning," in IEEE Access, vol.11, pp. 47410-47419, 2023, doi: 10.1109/ACCESS.2023.3274851.

[7] Zheng, Jiangbin, Yile Wang, Cheng Tan, Siyuan Li, Ge Wang, Jun Xia, Yidong Chen, and Stan Z. Li. "Cvt-slr: Contrastive visual-textual transformation for sign language recognition with variational alignment." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 23141-23150. 2023.

[8] Rajalakshmi, E., R. Elakkiya, V. Subramaniyaswamy, L. PrikhodkoAlexey, Grif Mikhail, Maxim Bakaev, Ketan Kotecha, Lubna Abdelkareim Gabralla, and Ajith Abraham. "Multi-semantic discriminative feature learning for sign gesture recognition using hybrid deep neural architecture." IEEE Access 11 (2023): 2226-2238.

[9] De Coster, Mathieu, Dimitar Shterionov, Mieke Van Herreweghe, and Joni Dambre. "Machine translation from signed to spoken languages: State of the art and challenges." Universal Access in the Information Society (2023).

[10] Gangrade, Jayesh, and Jyoti Bharti. "Vision-based hand gesture recognition for Indian sign language using convolution neural network." IETE Journal of Research 69.2 (2023): 723-732.

[11] Abualkishik, Abdallah, Wael Alzyadat, Marwan Al Share, Sara Al-Khaifi, and Mojtaba Nazari. "Intelligent Gesture Recognition System for Deaf People by CNN and IoT." International Journal of Advances in Soft Computing Its Applications 15, no. 3 (2023).

[12] Alaria, Satish Kumar, Ashish Raj, Vivek Sharma, and Vijay Kumar. "Simulation and analysis of hand gesture recognition for indian sign language using CNN." International Journal on Recent and Innovation Trends in Computing and Communication 10, no. 4 (2022): 10-14.

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/568





International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

#### Volume 5, Issue 3, June 2025



[13] Mohammedali, Atyaf Hekmat, Hawraa H. Abbas, and Haider Ismael Shahadi. "Real-time sign language recognition system." Int J Health Sci 6.S4 (2022): 10384-10407.

[14] A. Singh, A. Wadhawan, M. Rakhra, U. Mittal, A. A. Ahdal and S. K.Jha, "Indian Sign Language Recognition System for Dynamic Signs," 022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2022, pp. 1-6, doi: 10.1109/ICRITO56286.2022.9964940.

[15] Halder, Arpita, and Akshit Tayade. "Real-time vernacular sign language recognition using mediapipe and machine learning." Journal homepage:www. ijpr. com ISSN 2582 (2021): 7421.

[16] Patil, D. B., and G. D. Nagoshe. "A Real Time Visual-Audio Translator for Disabled People to Communicate Using a Human-Computer Interface System." Int. Res.

[17] Murali, Romala Sri Lakshmi, L. D. Ramayya, and V. Anil Santosh."Sign language recognition system using convolutional neural network and computer vision." International Journal of Engineering Innovations in Advanced Technology ISSN (2020): 2582-1431.

[18] Wadhawan, A., Kumar, P. "Deep learning-based sign language recognition system for static signs." Neural Comput and Applic 32, 7957–7968 (2020).

[19] Tolentino, L. K. S., Juan, R. S., Thio-ac, A. C., Pamahoy, M. A. B., Forteza, J. R. R., and Garcia, X. J. O. (2019). "Static sign language recognition using deep learning. International Journal of Machine Learning and Computing", 9(6), 821-827.

[20] Rokade, Y. I., and Jadav, P. M. (2017). "Indian sign language recognition system. International Journal of Engineering and Technology", 9(3), 189-196.

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/568

