Archives available at journals.mriindia.com

International Journal of Recent Advances in Engineering and Technology

ISSN: 2347 - 2812
Volume 14 Issue 01s, 2025

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

[1]Ms. P. B. Gholap, [2]Prathmesh Patil, [3]Shreya Pansare, [4]Mayur Bhagade
[1 2 3 4]AIDS Engineering Jaihind College of Engineering, Kuran, Pune, India
Email: [1]pallavidumbare@gmail.com, [2]ppatil3652@gmail.com, [3]pansareshreya8@gmail.com, [4]maauti.jcoe@gmail.com

Peer Review Information

Submission: 1 Sept 2025

Revision: 28 Sept 2025

Acceptance: 12 Oct 2025

Keywords

Sign language, Convolution neural Network, LSTM, Landmark.

Abstract

The people with speech disabilities usually face problems interacting with normal people, and there is a need to find away 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 onastandard 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.

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.Additionalcomponents, 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, enablingreal-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 mustbeabletoguess the sign from the frame that awebcam 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 model.

DATASET

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 to128 * 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 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 imageswere 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.

fig.1: The sample from dataset

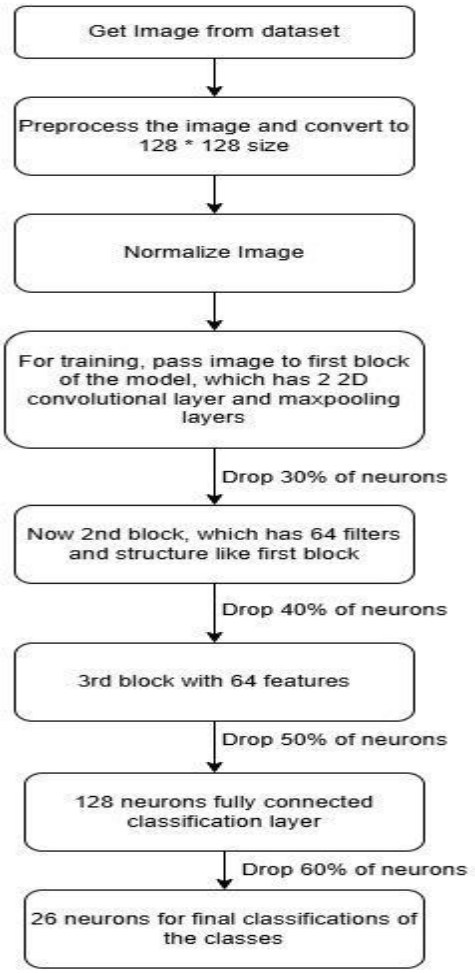
Methodology

The methodology for the project is to divide the recognition in the two parts. The first one is the Convolution 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.

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 asanimageand convert the image in the 128 * 128 size tomakean 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 averagepooling and fully connected classification layers. Each layer contains all the conv2D layers forfeature extraction. Every layer in the block has a Reluactivation and Batch regularization to train the model efficiently. The model faced some over fitting as some of the signs are similar to eachother.Such as the sign for letter D and P are somewhat similar so it occurred overfitting and was not able to trainas efficiently. To overcome this problem, we used L2 regularization. This approach reduced the possibility of overfitting. The kernel size set to3*3 with the same padding to preserve the spatial dimensions.

The first block has the 32 filters and two convolutional layers and a Max Pooling layer which removes all non-important features andconcentrates on only with higher values for further processing. To improve the generalization of the modelthedropoutlayer 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 ofparameters and mitigatetheoverfittingconcernsfor the model. There are fully connected classification layers total in 128. The dropout of 60% is introduced.



While training the model the stablelearning rateof0.0001isusedforstabletraining.Thisapproach 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.

Landmark Model

The second part is the landmarkmodelwhichplays one of the important roles in prediction of thesign. 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 Media pipe's hand landmark detection model returns 21 landmarks for one hand detectedintheframeandfortwohandsitreturns63 landmarks. The model is in such a way that it extracts landmarks from the frame if onlyonehand 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 has128fully connected classification layers which areconnected to 26 classification layers.

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.

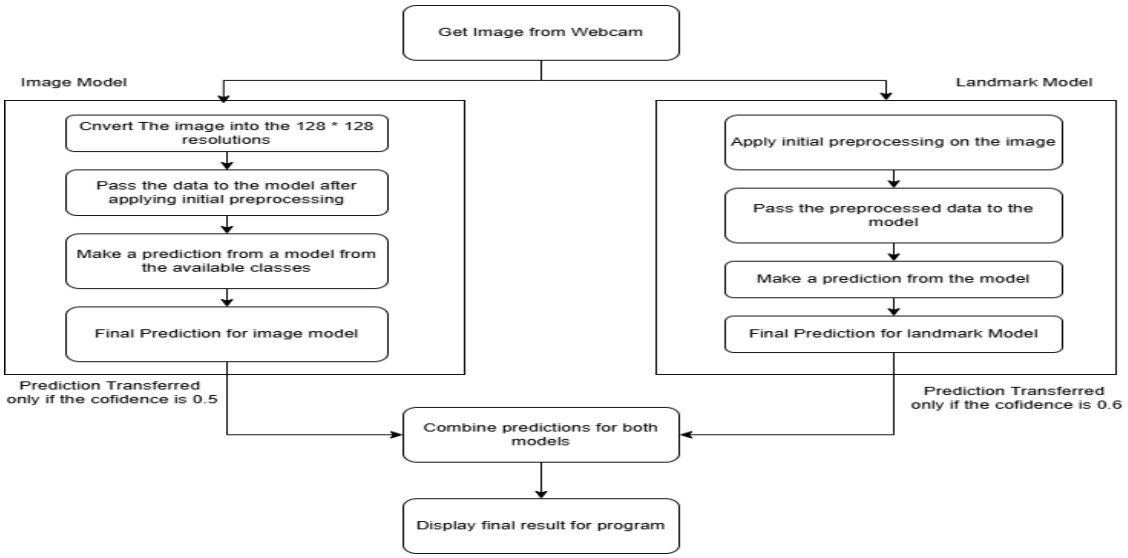


Fig. 2: Architecture of models and result finalization

RESULT

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 *128resolution.Themodelis based on convolutional neural networks (cnn) algorithms with conv2D layers in three different blocks. The model has around 0.9978 accuracyand 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 similarto 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.Thelandmarkmodelistrained on the

same dataset which was used to train the language model. The model was able tounderstand 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. Themodeltrainingaccuracywas97%andthevalue loss is around 0.4 but when treated withtheunseen 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 onlygathers 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.

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 sofor 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.

REFERENCES

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.

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

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.

Zhou, Benjia, Zhigang Chen, Albert Clap´es, Jun Wan, YanyanLiang,SergioEscalera, 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.

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.

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,"inIEEEAccess,vol.11,pp. 47410-47419,2023,doi: 10.1109/ACCESS.2023.3274851.

Zheng, Jiangbin, Yile Wang, Cheng Tan, Siyuan Li, Ge Wang, Jun Xia, Yidong Chen, andStanZ.Li."Cvt-slr:Contrastivevisual-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.

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

De Coster,Mathieu,DimitarShterionov,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).

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