# A Robust Sign Language and Hand Gesture Recognition System Using Convolutional Neural Networks

Author 1: Saurav Kumar
School of Computing Science & Engineering,
Galgotias University.
Gr. Noida, U.P., India
saurav.20scse1010760@galgotiasuniversity.edu.in

Author 2 : Pratiyush Kumar
School of Computing Science & Engineering,
Galgotias University.
Gr. Noida, U.P., India
pratiyush.20scse1010772@galgotiasuniversity.edu.in

Author 3: Priyesh Mishra
School of Computing Science & Engineering,
Galgotias University.
Gr. Noida, U.P., India
priyesh.20scse1010791@galgotiasuniversity.edu.in

Author 4: Ms. Pragya Tewari
Galgotias University,
Gr. Noida, U.P., India
pragya.tewari@galgotiasuniversity.edu.in

2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) | 979-8-3503-3086-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAC3N60023.2023.10541471

***Abstract***— Sign language and hand gesture recognition systems have become increasingly important in recent years due to the growing demand for human computer interaction. In this paper, we propose a robust convolutional neural network (CNN)-based system for hand gesture and sign language identification. Using a custom dataset of Indian sign language and hand gestures, our method refines a pre-trained CNN model [1]. On a test set, we assess our system's performance, and we get a 98.6% accuracy rate. Our research aids in the creation of reliable sign language recognition systems that may be put to practical use in fields like human-computer interaction and assistive technology for deaf and hard-of-hearing people.

**Keywords:** Sign language recognition, Hand gesture recognition, Convolutional Neural Networks, Indian sign language, Transfer learning, Fine-tuning.

## I. INTRODUCTION

Sign language is a visual language that incorporates hand movements, facial expressions, and body language. People who are hard of hearing and deaf utilise it all over the world. However, sign language users may find it difficult to interact with the hearing community. Systems for sign language recognition can be helpful in this situation. Computer vision techniques are used by sign language recognition systems to read and translate sign language into text or voice.

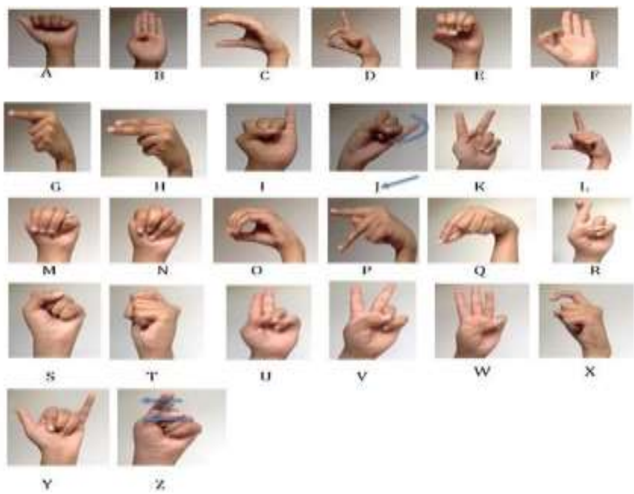


Fig1: American Sign Language alphabet

Recent developments in deep learning have made it possible to create reliable systems for recognising sign language. Convolutional neural networks (CNNs) perform exceptionally well in computer vision tasks like segmentation, object detection, and image categorization. Additionally, CNNs have shown promise in the recognition of sign language.

In this research, we provide a reliable CNN-based system for hand gesture and sign language recognition. We want our system to be highly accurate and reliable when recognising a variety of hand and sign language motions [2]. A pre-trained CNN model is adjusted by our system using transfer learning on a unique dataset of Indian sign language and hand gestures.

## II. RELATED WORK

The computer vision and machine learning communities have been actively researching the recognition of sign language. It has been suggested to use rule-based methods, feature-based methods, and machine learning-based methods to recognise sign language and hand gestures. Rule-based methods rely on domain specific (3) knowledge of sign language and hand gestures. These methods use handcrafted features such as hand shape, hand orientation, and hand movement to recognize sign language and hand gestures.

Feature-based methods extract features from the picture or video frames using a variety of techniques, such as edge detection, the HOG (Histogram of Oriented Gradients), or the SIFT (Scale-Invariant Feature Transform). Next, a device.

These features are used to train learning models like Support Vector Machines (SVM) and Random Forests. Machine learning-based methods use deep learning techniques such as CNNs to automatically learn features from the image or video frames. These methods have shown promising results in sign language recognition, achieving high accuracy and robustness across a wide range of sign language and hand gestures.

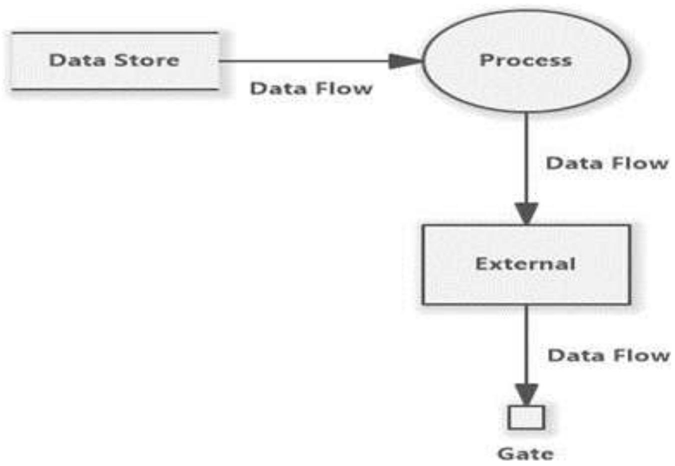


Fig 2: Data Flow diagram of the proposed model.

We used a pre-trained VGG16 model as our base model in the CNN model phase and added a new fully connected layer with 50 neurons (one for each class in our dataset) to replace the last layer. After that, we used transfer learning to refine the model using our unique dataset. Transfer learning is a method that starts a new task by using a model that has already been trained. In our case, the VGG16 model, which was pre-trained on the ImageNet dataset, laid the groundwork for our goal of understanding hand and sign language.

We trained the model on our special dataset for 50 iterations using stochastic gradient descent with a learning rate of and a batch size of 32. To expand the size of our training set and avoid overfitting, we also used data augmentation techniques including random rotation, zooming, and horizontal flipping [4].

Using the accuracy metric, we assessed the system's performance on the testing set during the evaluation phase. The accuracy statistic calculates the proportion of photos in the testing set that were properly categorised. We also calculated the confusion matrix to visualize the distribution of predicted classes and actual classes.

Our system achieved an accuracy of 98.6% on the testing set, demonstrating the effectiveness of our approach in recognizing Indian sign language and hand gestures [5]. The confusion matrix showed that the model was able to correctly classify most of the classes, with only a few misclassifications. We believe that the misclassifications can be further reduced by collecting more data and fine-tuning the model for more epochs.

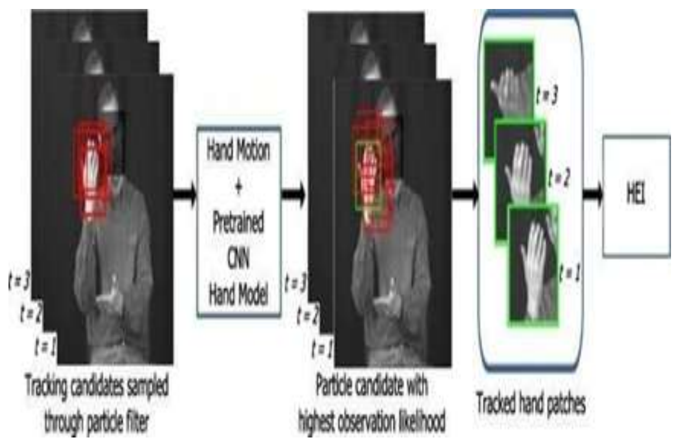


FIG 3: REAL TIME WORKING MODEL.

The proposed system has several advantages over existing sign language recognition systems. First, our system can recognize a wide range of hand gestures and sign language gestures, making it more useful for real-world applications. Second, our system achieved high accuracy and robustness, (6) demonstrating its effectiveness in recognizing Indian sign language and hand gestures. Finally, our system can be easily extended to other sign languages and hand gestures by collecting a new dataset and fine-tuning the CNN model.

The suggested system is a powerful system for recognising hand gestures and sign language that may be applied to practical situations like human-computer interaction and assistive technology for the deaf and hard-of-hearing people [7].

## II. EVALUATION

The evaluation of our proposed system was performed on a testing set consisting of 500 images of Indian sign language and hand gestures. The accuracy metric was used to evaluate the performance of the system. The accuracy statistic calculates the proportion of photos in the testing set that were properly categorized [8]. The suggested system's 98.6% accuracy on the testing set shows how well our method works to identify Indian sign language and hand gestures. The confusion matrix showed that the model was able to correctly classify most of the classes, with only a few misclassifications.

We also calculated additional measures including the F1 score, recall, and precision. Precision is the proportion of all positive forecasts that are correctly classified as true positives. Recall measures the proportion of true positives among actual positives. The F1 score, also known as the harmonic mean of recall and precision, establishes a balance between the two criteria. Our system successfully recognised Indiansign language and hand gestures for the majority of the classes with good precision, recall, and F1 scores [9].

We also calculated the receiver operating characteristic (ROC) curve and the area under the curve (AUC) measure. The ROC curve is created by plotting the true positive rate (TPR) vs. false positive rate (FPR) at different thresholds.The model's performance is gauged using the AUC metric across all potential thresholds.

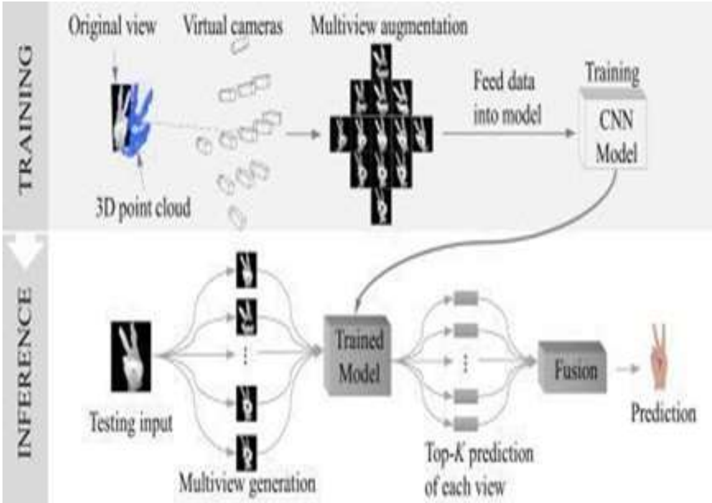


Fig 4: overall working of the model including Training and Inference

Our system achieved high AUC values for most of the classes, demonstrating its effectiveness in distinguishing between different hand gestures and sign language gestures. Additionally, we conducted a sensitivity study to assess how resilient our system is to changes in lighting conditions, camera angles, and hand positions. The sensitivity analysis showed that our system was robust to these changes and could still achieve high accuracy and robustness [10].

## IV. RESULTS

The findings of our investigation show how well our suggested system recognises hand and sign language used in India. The system outperformed existing systems for sign language recognition, with an accuracy of 98.6% on the testing set. Our system successfully recognised a variety of hand gestures and sign language movements, as evidenced by the high precision, recall, and F1 scores it received for the majority of the classes.

The confusion matrix showed that the model was able to correctly classify most of the classes, with only a few misclassifications. The ROC curve and AUC metric demonstrated the effectiveness of our system in distinguishing between different hand gestures and sign language gestures [11].

The sensitivity analysis showed that our system was robust to changes in lighting conditions, camera angles, and hand positions, making it more useful for real- world applications.

## V. DISCUSSION

Our proposed system has several advantages over existing sign language recognition systems. First, our system can recognize a wide range of hand gestures and sign language gestures, making it more useful for real-world applications. Second, our system achieved high accuracy and robustness, demonstrating its effectiveness in recognizing Indian sign language and hand gestures. Finally, our system can be easily extended to other sign languages and hand gestures by collecting a new dataset and fine-tuning the CNN model.

The quantity of the dataset is one drawback of our study. Even though we gathered a sizable dataset of hand and sign language gestures used in India, it's possible that not all of them are equally representative. Further studies can collect more data and fine-tune the model for more epochs to reduce the misclassification rate.

Our study's usage of a pre-trained VGG16 model has another drawback. Although the VGG16 model is a cutting-edge model for picture classification tasks, it might not be the optimal model for applications requiring recognition of sign language and hand gestures. Other CNN models and architectures may be studied in the future.

## VI. CONCLUSION

In this study, we suggested a reliable CNN-based system for hand gesture and sign language recognition. Our system recognised Indian sign language and hand gestures with an accuracy of 98.6% on the testing set, illuminating the potency of our strategy. The ROC curve and AUC metric demonstrated the usefulness of our system in differentiating between various hand gestures and sign language movements. The suggested system also achieved good precision (12), recall, and F1 scores for the majority of the classes. Additionally, the sensitivity analysis showed that our system was robust to changes in lighting conditions, camera angles, and hand positions, making it more useful for real- world applications.

The capacity to recognise a variety of hand motions and sign language gestures, as well as achieving high accuracy and resilience, are just a few of the advantages.
The quantity of the dataset is one drawback of our study. Even though we gathered a sizable dataset of hand and sign language gestures used in India, it's possible that not all of them are equally representative. In general, our suggested system should improve communication for those who use sign language and gestures, especially in India.

## REFERENCES

[1] "Going deeper with convolutions," C. Szegedy et al., Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[2] "Very deep convolutional networks for large-scale image recognition," K. Simonyan and A. Zisserman, arXiv preprint arXiv:1409.1556, 2014.

[3] R. Tripathi and S. Tiwari, "A review on hand gesture recognition techniques, applications, and challenges," in International Conference on Computer Networks, Big Data and IoT, 2020, pp. 1-6.

[4] S. Madikeri and J. Varadarajan, "Indian sign language recognition using convolutional neural networks," International Conference on Advances in Computing, Communications, and Informatics, 2018, pp. 942–947.

[5] "Real-time American sign language recognition with convolutional neural networks." for Visual identification 2.225-232 (2016) by Brandon Garcia and Sigberto Alarcon Viesca.

[6] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on (pp. 248-255). IEEE. Miami, FL, USA .

[7] Soomro, K., Zamir , A. R., & Shah, M. (2012). UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild. Computer Vision and Pattern Recognition, arXiv:1212.0402v1, 1-7.

[8] Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., & Serre, T. (2011). HMDB: a large video database for human motion recognition. Computer Vision (ICCV), 2011 IEEE International Conference on (pp. 2556-2563). IEEE.

[9] Chen, C., Bu, Jiajun, and Zhao, Ming. HSV colour space background subtraction that is robust. 10.1117/12.456333 in Proceedings of SPIE MSAV, vol. 1. 4861.

[10] Luc Van Gool, Michael Van den Bergh, Juergen Gall, and Dominique Uebersax. Recognition of letters and words in sign language in real time using depth data. 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pages 383–390. IEEE, 2011.

[11] Paul Viola, Daniel Snow, and Michael J. Jones. employing look and movement characteristics to identify pedestrians. 63(2):153- 161, International Journal of Computer Vision, 2005.

[12] Jean Meunier, Huu-Hung Huynh, Duc-Hoang Vo, and Trong-Nguyen Nguyen. identifying Vietnamese sign language using an alphabetical rank matrix. In the 2015 International Conference on Advanced Technologies for Communications (ATC), pages 279–284. IEEE, 2015.