



# 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 4, April 2025**

# Sign Language Interpretation using Convolution Neural Network

S. Jenita Christy, T. Anjani, A. Mani Deep, A. Nithin, A. Mounya Sri

Assistant professor, Dept. of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

B.Tech Students, Dept. of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

**ABSTRACT:** Sign language is a vital form of communication for the deaf and hard of hearing community. To bridge the communication gap between sign language users and non-signers, we present a novel system for real-time sign language detection using a standard web camera. This system aims to recognize sign language gestures performed in front of the camera, subsequently converting them into voice output and displaying the corresponding text on screen. The proposed system leverages computer vision techniques, including deep learning models, to capture and analyze sign language gestures. It first identifies the signer's hand and facial expressions, recognizing key markers that represent signs. The system then employs machine learning algorithms to translate these markers into sign language vocabulary. Upon successful detection and translation of the sign, the system provides simultaneous output in two ways: voice and on-screen text. The voice output enables real-time interpretation for users who may not be familiar with sign language, while the on-screen text serves as a visual reference. This dual output mechanism ensures accessibility and inclusivity for a wider audience. By integrating this system into webcams and other devices with cameras, we aim to enhance the communication capabilities of the deaf and hard of hearing community, enabling them to interact more effectively with hearing individuals. Additionally, this technology can find applications in education, healthcare, and other domains, fostering better understanding and accessibility for sign language users. **Key Words:** Sign language, Deaf and hard of hearing, Real-time, Sign language detection, Web camera, Computer vision

**KEYWORDS:** Image processing, CRNN, BERT Model, OCR, Cosine Similarities.

## I. INTRODUCTION

The project presented here is a groundbreaking endeavour that addresses a significant challenge faced by the deaf and hard of hearing community – the communication gap between sign language users and those who are not proficient in sign language. In this modern age where technology plays a pivotal role in enhancing our lives, a team of innovators has developed a novel system to bridge this gap using readily available tools – a standard web camera. Sign language is a vital and beautiful form of communication, allowing the deaf and hard of hearing to express themselves, connect with others, and access information. However, it remains a barrier for those who do not understand it, limiting inclusivity and access to vital services and information. This project aims to change that by introducing a real-time sign language detection system. The core of this system lies in cutting-edge computer vision techniques, particularly deep learning models, which enable it to capture and analyse sign language gestures accurately. When a signer interacts with the camera, the system identifies the signer's hand movements and facial expressions, recognizing key markers that represent specific signs. These markers are then translated into sign language vocabulary through the use of advanced machine learning algorithms. What makes this project truly exceptional is its dual output mechanism. Upon successful detection and translation of a sign, it provides simultaneous output in two ways: voice and on-screen text. The voice output enables real-time interpretation, making it accessible for individuals who may not be familiar with sign language. At the same time, the on-screen text serves as a visual reference, providing an additional layer of accessibility. This dual output mechanism ensures that a wider audience can engage and interact with sign language users fostering inclusivity and understanding. The potential impact of this project is substantial. By integrating this system into webcams and various devices with cameras, it has the power to enhance the communication capabilities of the deaf and hard of hearing community, allowing them to interact more effectively with the hearing world. Furthermore, this technology can extend its reach to multiple domains, such as education, healthcare, and many others, contributing to better understanding and accessibility for sign language users. In doing so, it represents a significant step forward in promoting inclusivity and empowering the deaf and hard of hearing community, showcasing the transformative power of technology in modern society.

## II.SYSTEM MODEL

The process through which the accuracy has been obtained is by using the dataset that has been described with an aim through which the approach move towards selecting the data, processing the data, training the data to selecting the appropriate model, which would give a good accuracy and remain suitable to the problem and extract the features from the given image in order to make a list to help the model, using activation and generalization for the model to properly respond and finally training the half dataset and the other half of the dataset for the testing purpose.

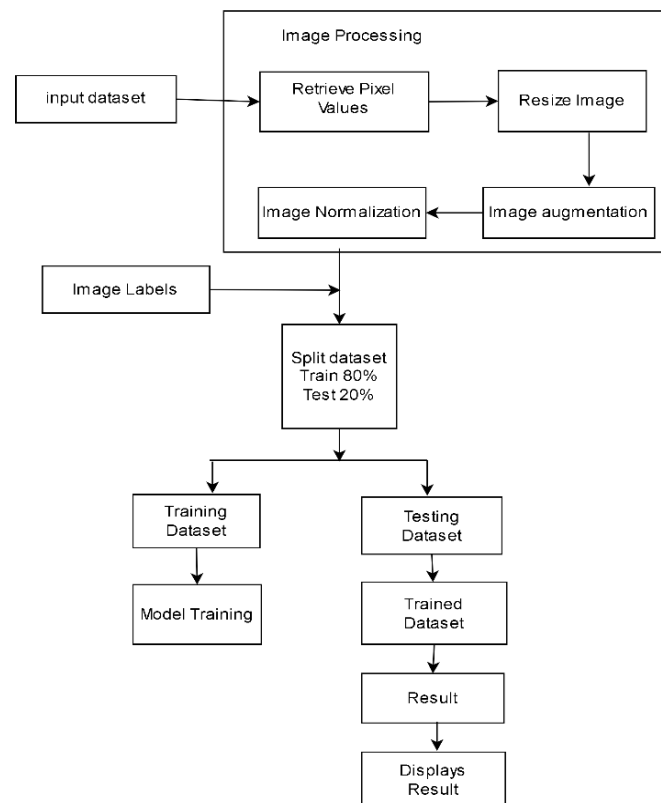


Fig 1: Design Architecture

The first thing to start with is to get the dataset for the model, you need to select an appropriate dataset which would be useful for the model to give a proper accuracy which can be also used in further different aspects. In order to perform data pre-processing using Python, we need to import some predefined Python libraries. These libraries are used to perform some specific jobs. Now the dataset selected should be imported into the model. After that we need to handle the missing data into the datasets. If our dataset contains some missing data, then it may create a huge problem for our machine learning model. Hence it is necessary to handle missing values present in the dataset. The dataset will be divided into training and validation collection in order to verify if the model we have taken is over fitted to the training dataset or not using the validation dataset. This helps us to train as well as test the same database so that we can know the proper accuracy of the model and also figure it out where it is going wrong and can easily correct it.

A Convolutional Neural Network is a class of deep neural networks, most commonly applied to analyzing visual imagery. They are also known as shift invariant or space invariant artificial neural networks, based on the shared weight architecture of the convolution kernels that scan the hidden layers and translation in variance characteristics. With most algorithms that handle image processing, the filters are typically created by an engineer based on heuristics. CNNs can learn what characteristics in the filters are the most important.



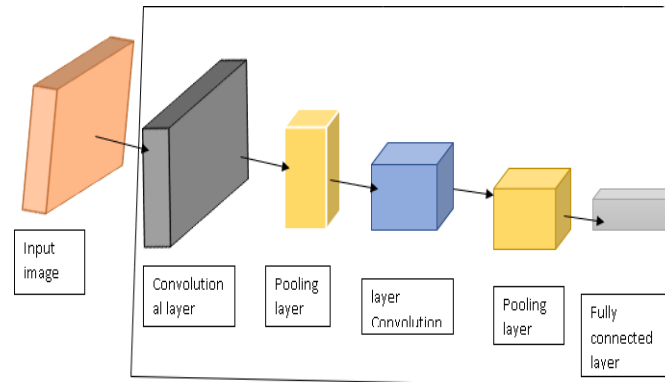


Fig:2

Education is nothing more than a learning cycle. In this section, we describe hyper parameters such as the number of epochs, batch size, learning rate, and so on. I used two callbacks one is 'early stopping' for avoiding overfitting training data and other 'ReduceLROnPlateau' for learning rate.

### III. LITERATURE REVIEW

Sandrine Tornay et al. [1], Sign language recognition involves modelling of multichannel information such as, hand shapes, hand movements. This requires also sufficient sign language specific data. This is a challenge as sign languages are inherently under-resourced. In the literature, it has been shown that hand shape information can be estimated by pooling resources from multiple sign languages. Such a capability. does not exist yet for modelling hand movement information. In this paper, we develop a multilingual sign language approach, where hand movement modeling is also done with target sign language independent data by derivation of hand movement subunits. We validate the proposed approach through an investigation on Swiss German Sign Language, German Sign Language and Turkish Sign Language, and demonstrate that sign language recognition systems can be effectively developed by using multilingual sign language resources. Hira Hameed et al. [2] Sign language is a means of communication between the deaf community and normal hearing people who use hand gestures, facial expressions, and body language to communicate. It has the same level of complexity as spoken language, but it does not employ the same sentence structure as English. The motions in sign language comprise a range of distinct hand and finger articulations that are occasionally synchronized with the head, face, and body. Existing sign language recognition systems are mainly camera-based, which have fundamental limitations of poor lighting conditions, potential training challenges with longer video sequence data, and serious privacy concerns. This study presents a first of its kind, contact-less and privacy-preserving British sign language (BSL) Recognition system using Radar and deep learning algorithms. Six most common emotions are considered in this proof-of-concept study, namely confused, depressed, happy, hate, lonely, and sad. The collected data is represented in the form of spectrograms. Three state-of-the-art deep learning models, namely, InceptionV3, VGG19, and VGG16 models then extract spatiotemporal features from the spectrogram. Finally, BSL emotions are accurately identified by classifying the spectrograms into considered emotion signs. Comparative simulation results demonstrate that a maximum classifying accuracy of 93.33 is obtained on all classes using the VGG16 model.

Aman Patel et al. [3] Sign language recognition plays an important role in real-time sign language translation, communication for deaf people, education and human-computer interaction. However, vision-based sign language recognition faces difficulties such as insufficient data, huge network models and poor timeliness. We use VTN (Video Transformer Net) to construct a lightweight sign language translation network. We construct the dataset called CSLBS (Chinese Sign Language-Bank and Station) and two-way VTN to train isolated sign language and compares it with I3D (Inflated three Dimension).

#### IV. RESULT AND DISCUSSION

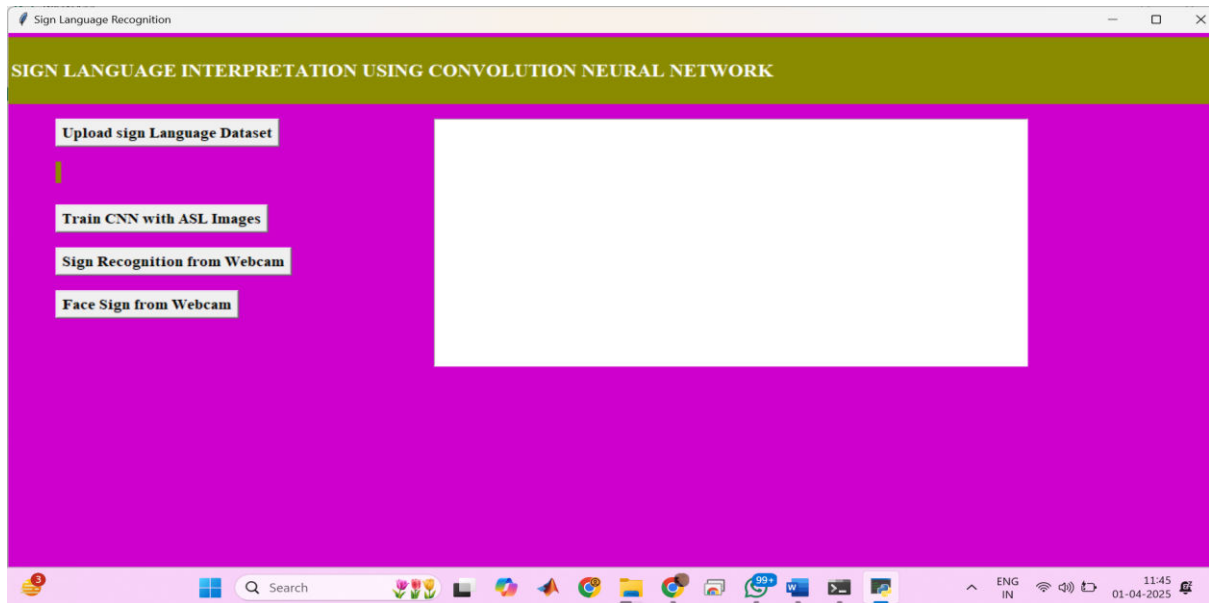


Fig:3

In above screen clicking on 'Upload Sign Language Dataset' button to upload dataset and then selecting and uploading Dataset folder to get below screen

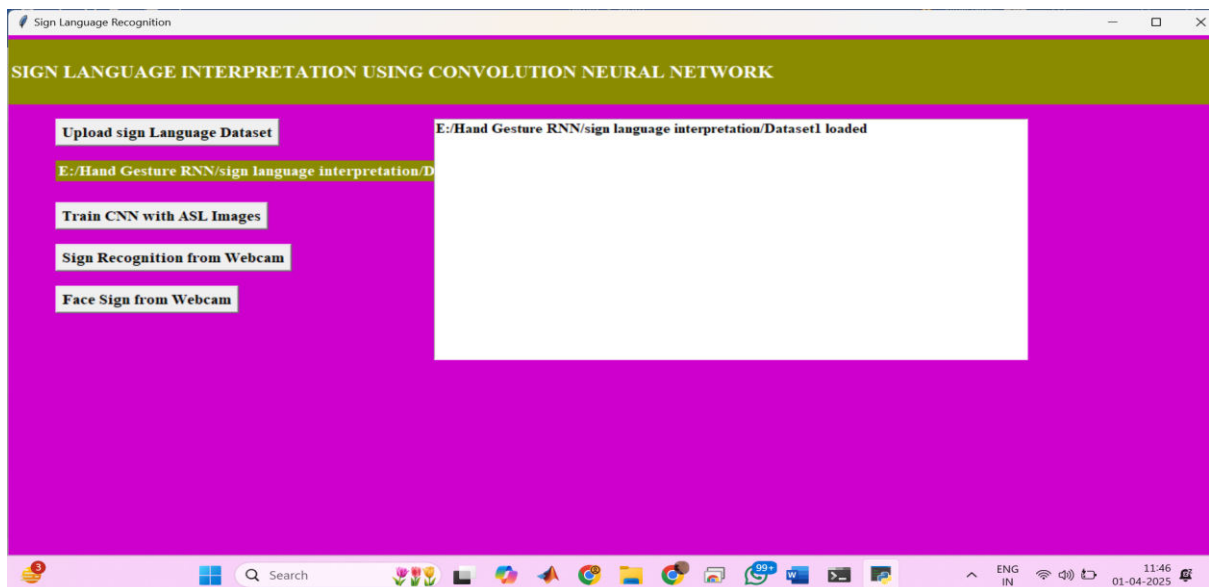


Fig:4

In above screen dataset loaded and now click on 'Train CNN with Sign Images' button to train CNN and get below output

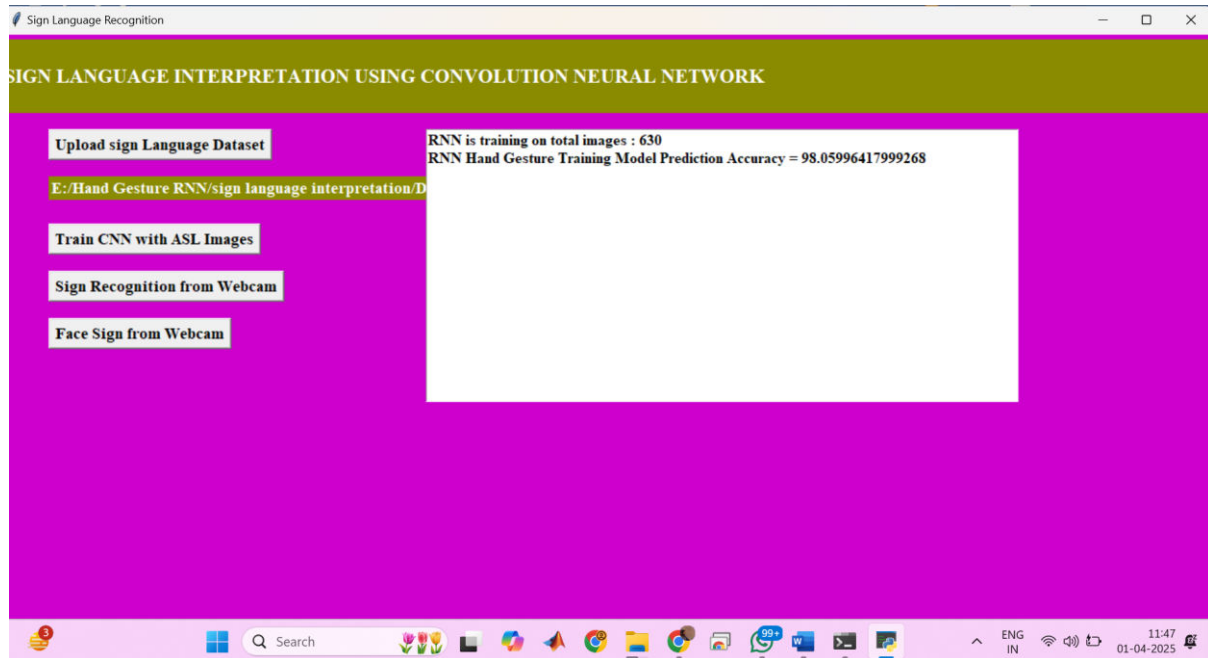


Fig:5

In above screen training is completed and we got CNN accuracy as 98% and now click on 'Hand Sign Recognition from webcam' button to start webcam and get below output

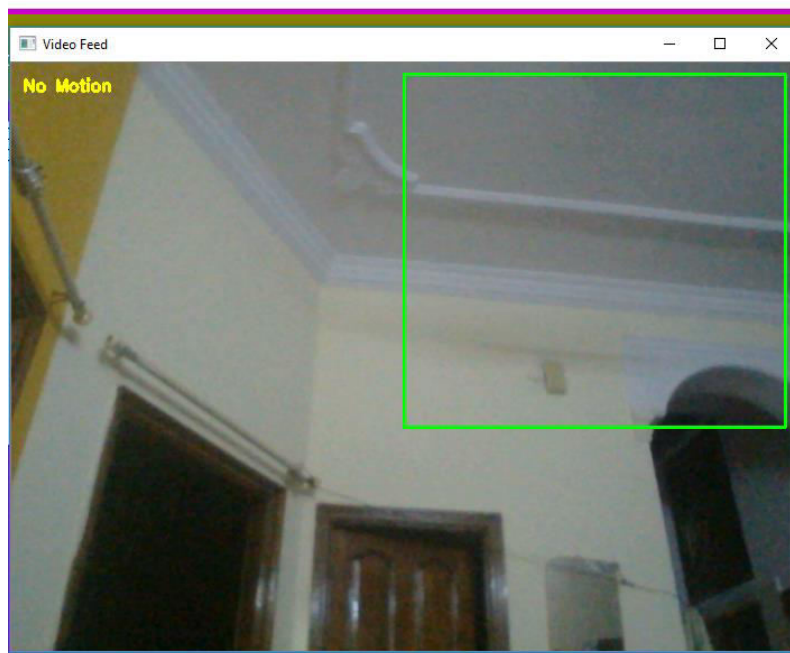


Fig:6

In above screen when no hand gesture visible then will get output as 'No Motion'.



Fig:7

In above screen we got output as Backward for shown gesture and you need to show hand gesture inside green colour rectangle box

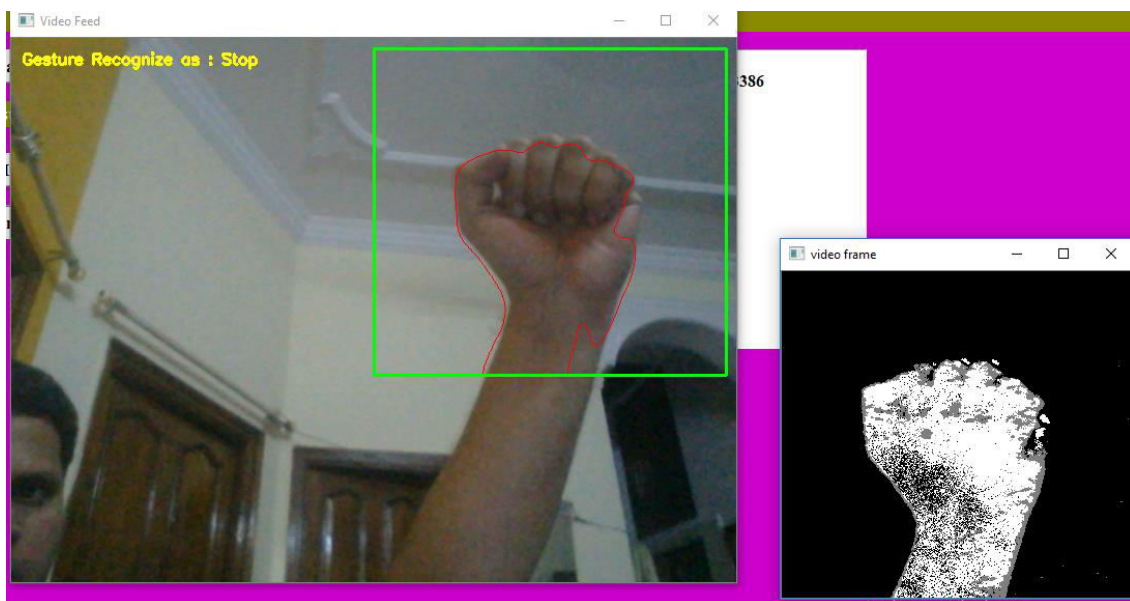


Fig:8

In above screen got output as Stop



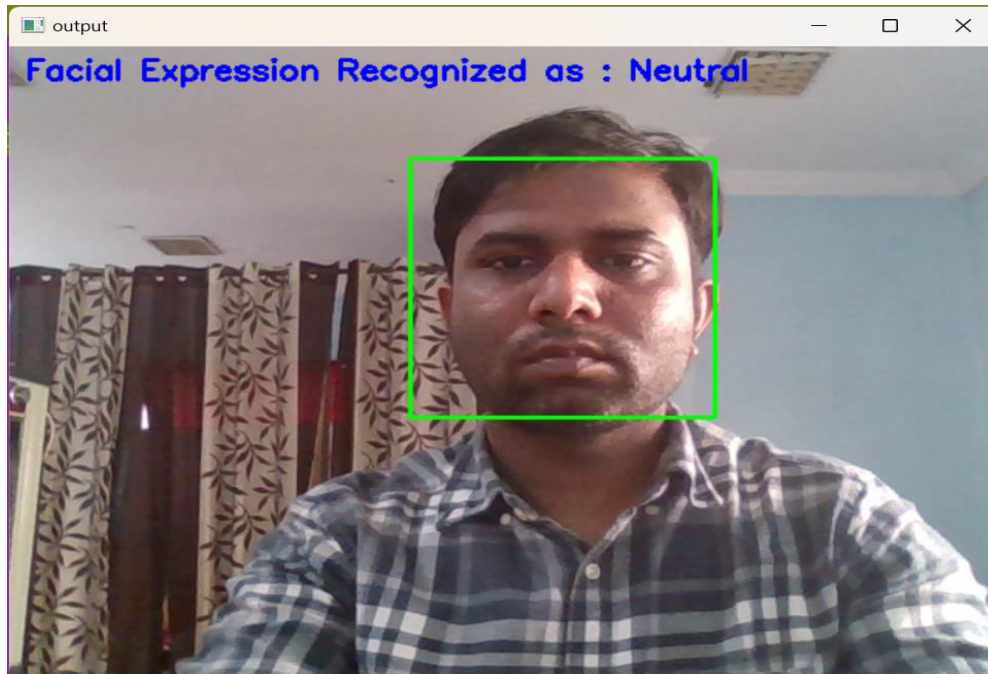


Fig:9

face sign identified as 'Neutral'

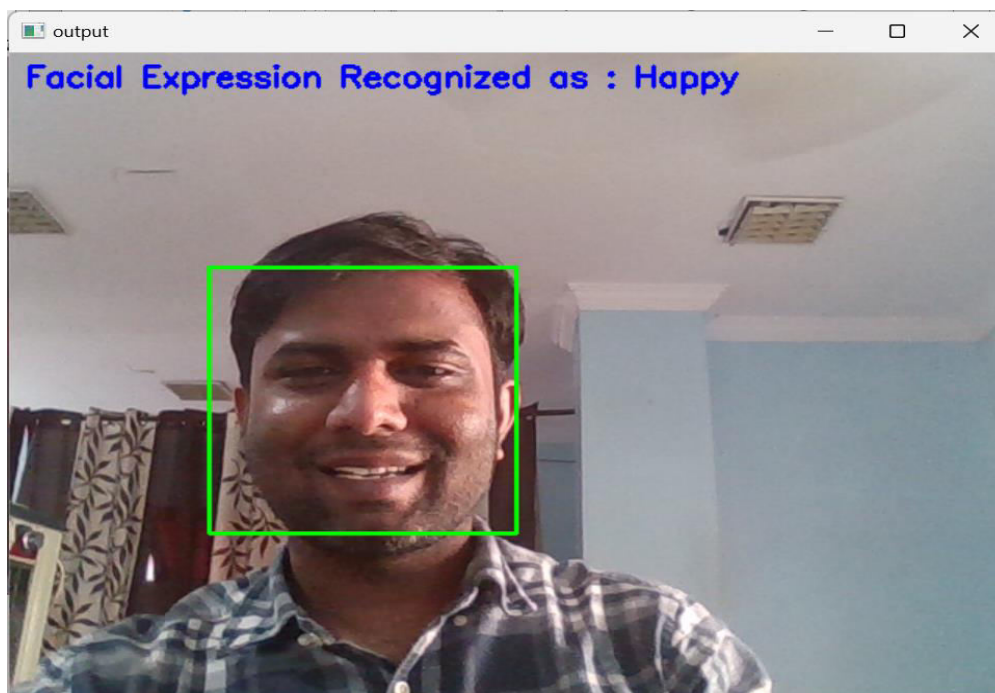


Fig:10

Face sign identified as 'Happy'



## VI. CONCLUSION

The proposed sign language interpretation system using CNN successfully bridges the communication gap between sign language users and non-signers. By leveraging deep learning and computer vision, the system accurately detects and translates sign language gestures into voice and text formats. The real-time processing capability enhances accessibility and usability across diverse environments such as education, healthcare, and public services. Compared to traditional interpretation methods, this approach offers an affordable and scalable solution without requiring specialized hardware. Future enhancements may include expanding the sign vocabulary, improving recognition accuracy under varying conditions, and integrating the system with mobile and wearable devices for enhanced accessibility. Overall, this project represents a significant step toward fostering better communication and inclusivity for the deaf and hard-of-hearing community.

## REFERENCES

- [1] Takalkar, Madhumita, Min Xu, Qiang Wu, and Zenon Chaczko. "A survey: facial micro- expression recognition." *Multimedia Tools and Applications* 77, no. 15 (2018): 19301-19325.
- [2] Adigun, Iyanu Pelumi, and Hima Bindu Vadapalli. "Facial micro-expression recognition: A machine learning approach." *Scientific African* 8 (2020): e00465.
- [3] Khor, Huai-Qian, John See, Raphael Chung Wei Phan, and Weiyao Lin. "Enriched long- term recurrent convolutional network for facial micro-expression recognition." In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pp. 667- 674. IEEE, 2018.
- [4] Li, Jing, Yandan Wang, John See, and Wenbin Liu. "Micro-expression recognition based on 3D flow convolutional neural network." *Pattern Analysis and Applications* 22 (2019): 1331- 1339.
- [5] Wang, Chongyang, Min Peng, Tao Bi, and Tong Chen. "Micro-attention for micro- expression recognition." *Neurocomputing* 410 (2020): 354-362.
- [6] Xia, Bin, Weikang Wang, Shangfei Wang, and Enhong Chen. "Learning from macro- expression: A micro-expression recognition framework." In *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 2936-2944. 2020.
- [7] Davison, Adrian K., Walied Merghani, and Moi Hoon Yap. "Objective classes for micro- facial expression recognition." *Journal of imaging* 4, no. 10 (2018): 119.
- [8] Xie, Hong-Xia, Ling Lo, Hong-Han Shuai, and Wen-Huang Cheng. "An overview of facial micro-expression analysis: Data, methodology and challenge." *IEEE Transactions on Affective Computing* 14, no. 3 (2022): 1857-1875.
- [9] Khor, Huai-Qian, John See, Sze-Teng Liong, Raphael CW Phan, and Weiyao Lin. "Dual- stream shallow networks for facial micro-expression recognition." In *2019 IEEE international conference on image processing (ICIP)*, pp. 36-40. IEEE, 2019.
- [10] Lei, Ling, Tong Chen, Shigang Li, and Jianfeng Li. "Micro-expression recognition based on facial graph representation learning and facial action unit fusion." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1571-1580. 2021.
- [17] Paul, Pias, Moh Anwar-Ul-Azim Bhuiya, Md Ayat Ullah, Molla Nazmus Saqib, Nabeel Mohammed, and Sifat Momen. "A modern approach for sign language interpretation using convolutional neural network." In *Pacific Rim International Conference on Artificial Intelligence*, pp. 431-444. Cham: Springer International Publishing, 2019.
- [18] Abiyev, Rahib H., Murat Arslan, and John Bush Idoko. "Sign language translation using deep convolutional neural networks." *KSII Transactions on Internet and Information Systems (TIIS)* 14, no. 2 (2020): 631-653.
- [19] Rao, G. Anantha, K. Syamala, P. V. V. Kishore, and A. S. C. S. Sastry. "Deep convolutional neural networks for sign language recognition." In *2018 conference on signal processing and communication engineering systems (SPACES)*, pp. 194-197. IEEE, 2018.
- [20] Garcia, Brandon, and Sigberto Alarcon Viesca. "Real-time American sign language recognition with convolutional neural networks." *Convolutional Neural Networks for Visual Recognition* 2, no. 225-232 (2016): 8.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



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

Scan to save the contact details