WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Audio-To-Sign Conversion and Hand Gesture Recognition with An Air Board for Deaf and Dumb Using Deep Learning

[1]Kaniti Kavya Sushma Mounika, [2]Juttuka Satya Hemalatha, [3]Ghantasala Divya Sri, [4]Mylavarapu Kalyan Ram

[1, 2, 3] VIII Semester B. Tech Students, [4]Associate Professor

[1,2,3,4]Department of CSE, Aditya Engineering College, Surampalem, A.P., India

ABSTRACT:

Every Language has syntax and set of rules for connecting words to make statements. In Sign Language, also, there are different signs used by deaf and dumb people for communicating with others. To make a conversation with the deaf and dumb people, we have to learn sign language which is a difficult task. If a person is speaking with a hearing-impaired person, he may or may not understand the speaker and the speaker also cannot understand the sign language of the hearing-impaired person. So, it is necessary to learn the Sign Language if a person wants to make an understandable conversation with the deaf and dumb people. This proposed model is a desktop-based app, designed and developed using Python programming language, and the technology used is Deep learning. Convolutional Neural Network (CNN) is a Deep Learning Technique used for analysing the camera feed and to detect the signs. This model is capable of taking inputs in both image and speech format and can convert the hand gestures into text form as well as the audio into sign language. Also, can recognize the letter written in the air. The output of this model is displayed on the screen of the user's desktop in the form of text and images. The main aim of the proposed model is to bridge the communication gap between common people and the deaf and dumb people.

Key Words– Deep Learning, Convolutional Neural Network (CNN), Audio to Sign Conversion, Hand Gesture Recognition, Sign to text, Air Board, Write on Air.

INTRODUCTION

Without learning sign language, communicating with deaf and dumb persons is a difficult task. It cannot be easily understandable by common people, and the hearing impaired cannot understand what the common people are conveying. Therefore, here there occurs communication problem between them. This project aims to lowers the communication gap between this community and additionally the standard world.

This paper proposed a model that makes ease of communication between the common people and the hearing-impaired people. It is a Desktop based application which is designed using Python programming language and uses Deep Learning technique for data analysis known as Convolutional Neural Network (CNN). In this, sequential 2D CNN model is used for data analysis and to classify the images that are captured from live video frames. This model contains three main features. They are:

(i). Audio to Sign Conversion, uses a python library (SpeechRecognition) & converts audio into text. Next, the text is converted into Sign language.

(ii). Hand Gesture Recognition, which Convolutional Neural Network model (2D) for Recognizing Gestures.

(iii). Air Board, which uses camera feed to process when the user writes on air using a blue object then predict the letter written by user on air using a blue object.

RELATED WORKS

In this paper, G. Park, V. K. Chandrasegar and J. Koh, introduced a model which aims to have better human device interaction. This model recognizes hand gesture of user using Recurrent Neural Network (RNN). It requires a Sinuous antenna to collect data from the five hand signals and to conditioning them, it uses RNN. This model cannot able to convert audio into the sign language and Air Board feature. [1]

In this paper, S. Pariselvam, D. N., D. S. and S. B., proposed a model that provides Human Computer Interaction using CNN model. This model proposed two features which are mainly used for abnormal people. They are Audio to Sign Conversion and Hand Gesture Recognition. This model doesn't have Air Board feature included in it.[2]

In this paper, Z. Wang et al., proposed a model that is a device-free hand gesture recognition system based on Channel State Information (CSI) and deep learning models, called WiDG. This system is limited to recognize the handwritten digits from 0 to 9 in the air and it cannot convert speech into sign language and there is no air board feature in it. [3]

In this paper, A. Yadav, R. Saxena, B. Saini, V. K. Verma and V. Srivastava, proposed a model that is a web-based application which can able to convert speech to Signs. If the related words are not found in the database, then it looks for its related synonyms and replaces with term. It does not support the conversion of hand gesture into text form and don't have Write on Air feature. [4]

In this paper, T. A. Siby, S. Pal, J. Arlina and S. Nagaraju, proposed a model that uses CNN to recognize the hand gestures made by the user through a web camera. If the gesture is recognized, then the text is displayed on the output screen. Using gTTS library, it can convert that text into audio. This model does not support Write on Air feature. [5]

In this paper, S. Gupta, R. Thakur, V. Maheshwari and N. Pulgam, proposed a model that is to build a system which helps the paralyzed people to communicate with others through Hand Gestures. It can convert signs in to text. This system can alert the people, in case, there is an emergency. [6]

In this paper, Grigoris Bastas, Kosmas Kritsis, Vassilis Katsouros, proposed a model which supports air writing recognition feature. It can track the fingertip written digits on air from 0 to 9 using deep CNN architecture. This model doesn't have the Audio to sign conversion and sign detection features. [7]

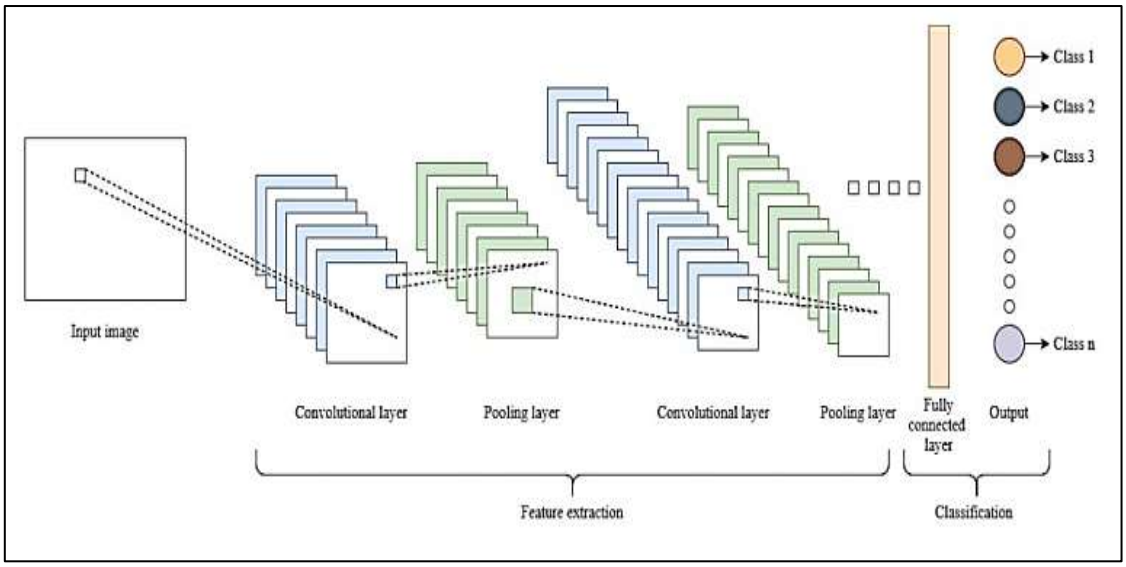
In this paper, Chaur-Heh Hsieh, You-Shen Lo, Jen-Yang Chen, Sheng-Kai Tang, proposed a model that enables user to write on air using deep Convolutional Neural Network (2D CNN model). This model can only able to recognize numerical digits that are written on air using fingertip and doesn't support Audio to Sign and Sign to text features.[8]

In this paper, Muhammad Arsalan, Avik Santra, proposed a model that is a virtual board is used for a network of 60-GHz milli-meter wave frequency modulated continuous wave (FMCW) radars. It detects the trajectory of hand marker writing motion. It needs a radar set up to run the model. It doesn't support Audio to sign language conversion and sign detection features. [9]

In this paper, Seong Kyu Leem, Faheem Khan, Sung Ho Cho, proposed a model that classifies fingertip digits written in mid-air. This model uses CNN to classify the image which contains trajectory data. To, identify the trajectory of the data written in air, Impulse radio ultrawideband (IR-UWB) radar sensors setup is needed. It doesn't have the Audio to sign language conversion and sign detection features. [10]

METHODOLOGY

The proposed model uses Convolutional Neural Networks (CNN) a deep learning technique to recognize the hand gestures made by the user, and identification of the Hindi characters that are written on air using blue object. The CNN model is a neural network that contains layers which takes exactly one input and gives one output. These layers are grouped together and forms a network. The CNN model is mostly used for analysing the images and videos because it gives more accuracy in classification. The below shown figure is the architecture of CNN model.



Architecture of CNN Model

There are two main parts in the architecture of CNN model. They are Feature Extraction and Classification.

- Feature Extraction is a convolution tool which is responsible for separating, identifying and analysing the features of the image. It includes Convolutional layer which is responsible for feature detection and Max pooling layer which is responsible for feature selection.

- Classification layer is responsible for classifying the images. Classification layer includes fully connected layer. In fully connected layer, all nodes are connected with one another. The input for this layer is the output of Feature Extraction layer.

PROPOSED MODEL

The main aim of the proposed model is to make ease of communication in an understandable and an efficient way between hearing impaired and common people. This proposed model contains three features. These three features help to make an effective communication between hearing impaired people and common people. They are:

- **Audio to Sign Conversion**:

This module is responsible for converting audio into sign language. In this, the input is audio/speech. When the user speaks, this model recognizes it and converts it into text using Speech_recognition library. Later, the text classified and is converted into Sign language.

- **Hand Gesture Recognition**:

This module is responsible for converting sign to text or we can say it detects the sign shown in front of the camera using CNN 2D model. In this, the 2D CNN model is trained and tested with the hand gestures of English alphabet dataset.

- **Air Board**:

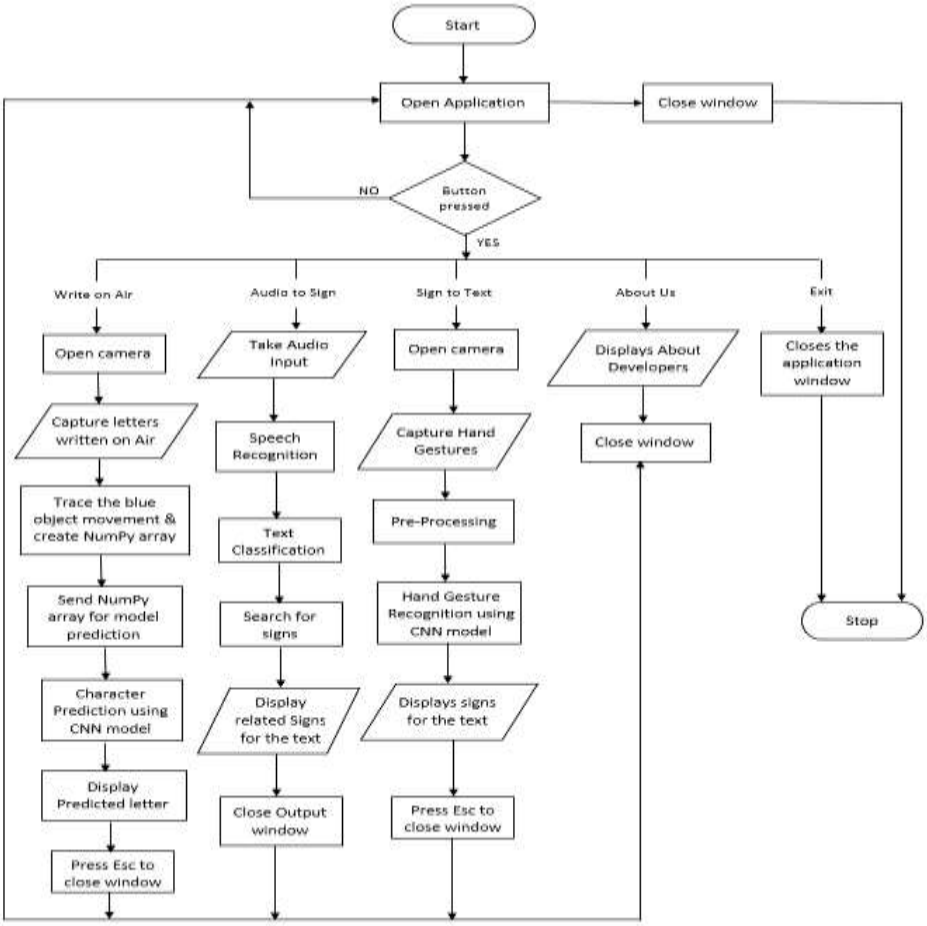
This feature is responsible for tracking the blue object movement when the user writes on air using blue object and recognizes Hindi characters that are written on air. In this, the CNN model is trained and tested with the Hindi hand written letters dataset which helps to recognizes the Hindi letters that are written on air.

Advantages:

- The main advantage of this model, the three features included in it which makes it more efficient than other existing systems.
- It helps to communicate and understand the deaf and dumb people in an efficient way.
- The users can communicate and also learn sign language using the audio to sign conversion.
- It is more efficient due to the features included in it.
- Average accuracy of the proposed model is more when compared to the other models.
- It is a user-friendly application and free of cost.

FLOWCHART DIAGRAM

The following shown diagram is the flowchart diagram of the project which shows the sequence of actions involved in execution of the proposed model:

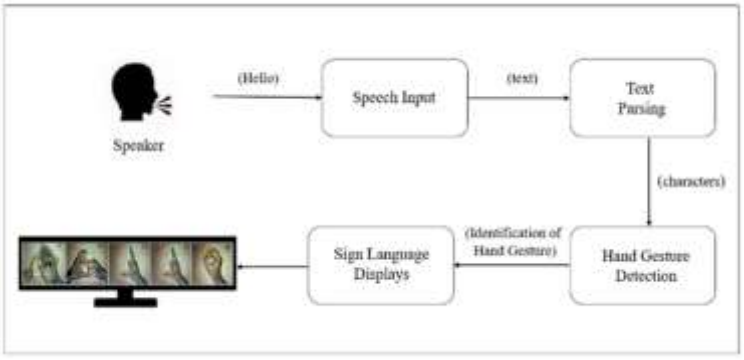


SYSTEM IMPLEMENTATION

The proposed system is implemented using Python programming language in Visual code Studio IDE. To build the sequential two-dimensional Convolutional Neural Network model (CNN), Python programming language is most commonly used. The CNN model can be built easily using model in Keras, an API used for providing high level abstraction to build the CNN model and TensorFlow framework used for training the CNN model layer by layer. In the proposed model, the sequential 2D CNN model is used for converting the sign language into text and to recognize the Hindi characters that are written on air using blue object. The features proposed in this system are designed in such a way that each feature performs a different task and has different workflows. The project workflows of each feature are explained below.

1) Project flow of Audio-To-Sign Conversion:

Whenever user speaks “Hello Everyone”, using speech_recognition library the speech is recognized and converted to text then that text. Next, it will be split into words then into the characters using python programming. Next, the respective hand gesture for the given input is displayed in a window on the desktop of the users in a series of signs images as shown in the figure. If the input speech is not clear, then the output will be NONE which will be displayed in sign language only.



Project flow of Audio to Sign Conversion

Implementation:

1. First of all, download the hand gestures of English alphabets from Google.
2. Store them in a folder and save it with the respective English alphabet.
3. Now the dataset is ready.
4. To take the input from the user, the model asks to give input. (“Speak any word, that you want to convert in sign language”)
5. To recognize the audio input, SpeechRecognition module in Python, a Google API is used to recognize the speech and converts it into text.
6. Next, this text is split into words using a built-in function of python split().
7. For example: “Hello World” = {‘hello’, ‘world’}
8. Again, these words are split into letters by looping each word.
9. For example:

{‘hello’, ‘world’} = {‘h’, ‘e’, ‘l’, ‘l’, ‘o’, ‘w’, ‘o’, ‘r’, ‘l’, ‘d’}

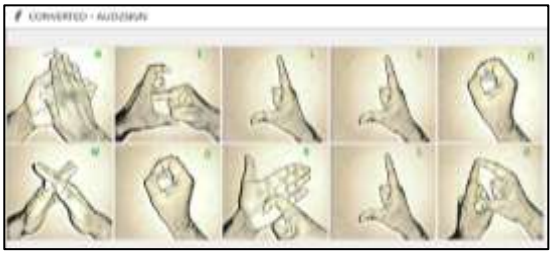
10. Next, these letters are looped again and are searched in the folder where the signs are stored.
11. If the sign is found, then that image is added to the output frame, otherwise, the model says “Sorry! This System accepts only alphabets” and “Can't be Converted to sign language”.
12. If the Audio input is not clear or didn’t understand by the model then it will display the word ‘NONE’ in the sign language (in image format).
13. The final output consists of the series of related signs with respective to the letters present in the Audio.
14. And the output will be in the word after word format.

Tests and Results

- **Testcase1:** Sign Conversion of audio: “Hello World”

Input: “Hello World”

Output:



Status: Pass

- **Testcase2:** If audio is not recognized

Input: unclear

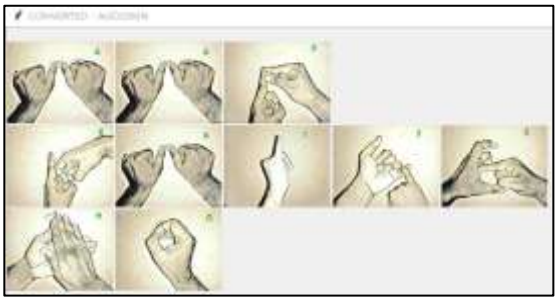
Output:

Status: Pass

- **Testcase3:** Sign Conversion of audio “aap kaise ho”

Input: “aap kaise ho”

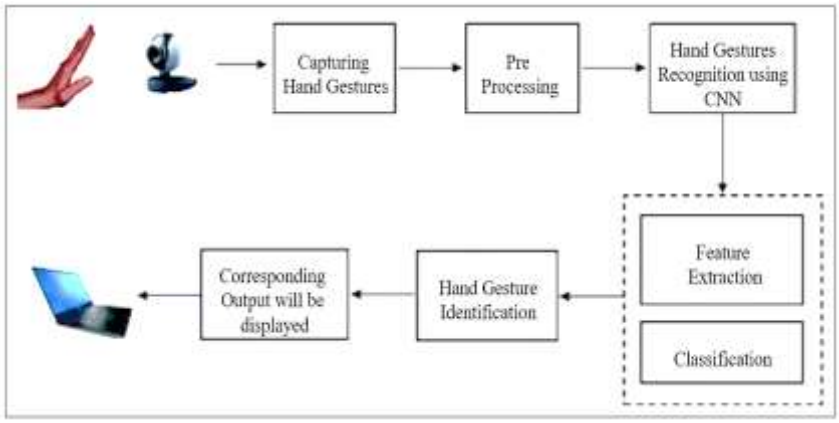
Output:



Status: Pass

2) Project flow of the Hand Gesture Recognition:

Whenever user shows some signs in front of the camera as shown in the figure then it captures the hand gestures made by the users. Next, it does the pre-processing and eliminates background noises then recognizes the hand gestures using 2D CNN model by feature extraction and classification. After identification of Hand Gesture, the output will be displayed on the screen of the desktop to the user through text message.



Project flow of Hand Gesture Recognition

Implementation:

1. The first step to design the Hand Gesture Recognition system, is to download the dataset from UCI Machine Learning Repository. Next, to capture live frames from webcam, OpenCV module is loaded.
2. For training, CNN2D sequential model is used. For that dataset need to be shuffled and then divided to 80-20 ratio.
3. Now, the layers of 2D CNN model are trained with the dataset.
4. Layers of sequential model are: CONV2D > AVERAGEPOOLING2D > DROPOUT > CONV2D > AVERAGEPOOLING2D > DROPOUT > FLATTEN > DENSE > DROPOUT > DENSE.
5. Use the activation function as ReLu.
6. The training data is now passed through these layers to fit the model.
7. Next, the model is tested with training data to evaluate the model.
8. Now save the model.
9. Load the model.
10. Setting a square box in the camera frame for detecting the hand gestures.
11. Applying flip, cvtColor of OpenCV on the camera frame for removing noise and detecting the hand gestures.
12. Once the Hand Gesture is found then that gesture is sent to the CNN model to predict the character.
13. Before prediction, the captured images need to be pre-processed by resizing it, converting to NumPy array, and reshaping it.
14. This array is used as parameter for keras function "predict".
15. Predict function predicts some value and that value is searched in dictionary of characters.
16. If the value is found, then that character is printed on the screen.

Tests and Results

- **Testcase:** Detection of Hand Gesture 'f'

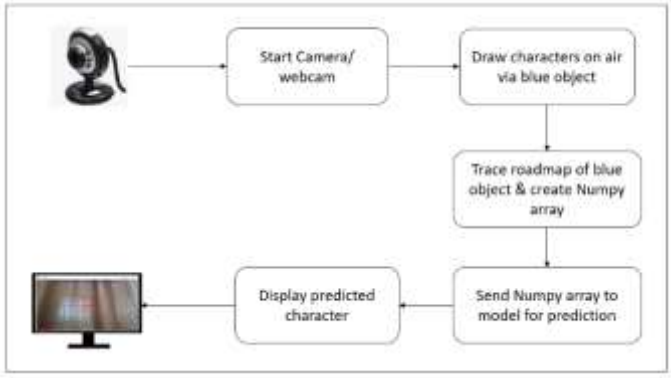
Input:

Output: f

Status: Pass

3) Project flow of the Air Board:

Whenever user tries to write on air using a blue object in front of the camera, then the model uses a python module OpenCV to capture the live frames from webcam. When the user moves a blue object in front of the camera, it detects the blue object and track the movement of the blue object then draws a line in the route of movement. This model then detects the Hindi letter that is written on air. Using Convolutional Neural Network (CNN 2D model), the value is searched in dictionary of characters. If found then it displays the detected value on the camera frame itself.



Project Flow of the Air Board

Implementation:

1. First of all, we need to download the dataset from UCI Machine Learning Repository.

(https://archive.ics.uci.edu/ml/datasets/Devanagari+Handwritten+Character+Dataset#)

2. Dataset includes .png format images of resolution 32*32, so we need to convert the dataset to csv file.
3. For training, CNN2D sequential model is used. For that dataset is divided into 80:20 ratio for training and testing purpose.
4. Prepare the architecture of 2D CNN model.

Layers of sequential model are: CONV2D > AVERAGEPOOLING2D > DROPOUT > CONV2D > AVERAGEPOOLING2D > DROPOUT > FLATTEN > DENSE > DROPOUT > DENSE.

5. Use the activation function as ReLu.
6. This model is passed through all these layers to fit the training data and set epochs as 35. with batch size of 64.
7. After finishing it, we will send testing data to evaluate the testing, then save the model.
8. Load the model and capture live frames from webcam.
9. Setting the upper and lower range of blue colour, for detecting the blue colour object.

10. To remove background noises and detecting the blue coloured object, apply flip, cvtColor, inRange, medianBlur, GaussianBlur & threshold layers of OpenCV into frame.
11. Track the movement of the blue object and draw line in the route of movement
12. Once the blue object is found, sending the gesture motion to predict the character.
17. Before prediction, the captured images need to be pre-processed by resizing it, converting to NumPy array, and reshaping it.
18. This array is used as parameter for keras function "predict".
19. Predict function gives some value between 0 to 37 and that value is searched against the dataset.
20. If the value is found, then that character is printed on the screen.

Tests and Results

- **Testcase1:** Blue Object Identification

Status: Pass

- **Testcase 2:** Recognition of Hindi letter written on air ‘ra’

Input:

Output:

Status: Pass

CONCLUSION

Without learning sign language, communicating with deaf and dumb persons is a difficult task. It cannot be easily understandable by common people, and the hearing impaired cannot understand what the common people are conveying. Therefore, here there occurs communication problem between them. Using, this model, the communication gap between hearing disabled person and the common people.

There are mainly three features include in this proposed model which are Audio to Sign Language Conversion, Hand Gesture Recognition, and Write on Air/Air Board which helps to communicate with hearing disabled persons in a better way. The average accuracy of the entire project is more when compared to other models. This project includes improving accuracy and expanding the vocabulary of recognized signs to increase its usability and effectiveness for users. This proposed model can able to provide an efficient and intuitive way to communicate in sign language.

FUTURE ENHANCEMENT

In future, new features can be included in this model to make an efficient communication with non-hearing individuals. In Hand gestures recognition could be used in security applications to identify individuals based on their unique hand gestures. To create more comprehensive and intelligent system for communication, we could expand the project to recognize letters for multiple languages. It increases its usability for wider audience and to perform tasks by writing commands in air. The Hand Gestures Recognition can be expanded by recognizing the hand gestures even in noisy background.

REFERENCES

[1] G. Park, V. K. Chandrasegar and J. Koh, "Hand Gesture Recognition using Deep learning Method," 2021 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI), 2021, pp. 1347-1348, doi: 10.1109/APS/URSI47566.2021.9703901.

[2] S. Pariselvam, D. N., D. S. and S. B., "An Interaction System Using Speech and Gesture Based on CNN," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1-5, doi: 10.1109/ICSCAN49426.2020.9262343.

[3] Z. Wang et al., "WiDG: An Air Hand Gesture Recognition System Based on CSI and Deep Learning," 2021 33rd Chinese Control and Decision Conference (CCDC), 2021, pp. 1243-1248, doi: 10.1109/CCDC52312.2021.9602438.

[4] A. Yadav, R. Saxena, B. Saini, V. K. Verma and V. Srivastava, "Audio to Sign Language Translator Web Application," 2021 International Conference on Computational Performance Evaluation (ComPE), 2021, pp. 321-326, doi: 10.1109/ComPE53109.2021.9751857.

[5] T. A. Siby, S. Pal, J. Arlina and S. Nagaraju, "Gesture based Real-Time Sign Language Recognition System," 2022 International Conference on Connected Systems & Intelligence (CSI), 2022, pp. 1-6, doi: 10.1109/CSI54720.2022.9924024.

[6] S. Gupta, R. Thakur, V. Maheshwari and N. Pulgam, "Sign Language Converter Using Hand Gestures," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. 251-256, doi: 10.1109/ICISS49785.2020.9315964.

[7] G. Bastas, K. Kritsis and V. Katsouros, "Air-Writing Recognition using Deep Convolutional and Recurrent Neural Network Architectures," 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR), Dortmund, Germany, 2020, pp. 7-12, doi: 10.1109/ICFHR2020.2020.00013.

[8] Hsieh, Chaur-Heh & Lo, You-Shen & Chen, Jen-Yang & Tang, Sheng-Kai. (2021). Air-Writing Recognition Based on Deep Convolutional Neural Networks. IEEE Access. 9. 142827-142836. 10.1109/ACCESS.2021.3121093.

[9] M. Arsalan and A. Santra, "Character Recognition in Air-Writing Based on Network of Radars for Human-Machine Interface," in IEEE Sensors Journal, vol. 19, no. 19, pp. 8855-8864, 1 Oct.1, 2019, doi: 10.1109/JSEN.2019.2922395.

[10] S. K. Leem, F. Khan and S. H. Cho, "Detecting Mid-Air Gestures for Digit Writing With Radio Sensors and a CNN," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 4, pp. 1066-1081, April 2020, doi: 10.1109/TIM.2019.2909249.