e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

( Peer-Reviewed, Open Access, Fully Refereed International Journal )

Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

# DYNAMIC SIGN LANGUAGE INTERPRETATION PLATFORM

**Godwin Immanuel Kingston. J[*1], Dr. Ilamchezhian. J[*2], Sarala Devi. V[*3]**

[*1]MCA Student, Department Of Computer Applications Dr. M.G.R Educational And Research Institute, Chennai, India.

[*2,3]Asst. Professor, Department Of Computer Applications Dr. M.G.R Educational And Research Institute, Chennai, India.

## ABSTRACT

Hearing and speech impairments affect over 36 million individuals globally, highlighting the critical need for advancements in communication accessibility. Addressing this need, our project focuses on developing a real-time sign language recognition system leveraging deep learning techniques. The system aims to interpret American Sign Language (ASL) alphabet gestures captured through image and video datasets, providing instantaneous output in either text or audio format based on user preference. Central to its functionality is the accurate extraction of hand gestures facilitated by integrated sensing devices. This ensures effective communication between individuals with and without hearing impairments.

Furthermore, our system's versatility extends its utility beyond individual communication needs, presenting significant potential for educational applications. By enhancing accessibility and inclusivity, it can empower students with hearing impairments to engage more effectively in educational settings. Educators can leverage this technology to create inclusive learning environments and promote equal opportunities for all students.

In summary, our real-time sign language recognition system represents a pivotal advancement in addressing communication barriers faced by individuals with hearing impairments. Through the integration of deep learning algorithms and advanced sensing technologies, it offers a comprehensive solution for facilitating seamless communication and promoting inclusivity within individual and educational contexts.

**Keywords:** Sign Language Recognition, Real-Time Gesture Recognition Sign Language (ASL), Hand Gesture Detection Hand Posture Recognition, Hand Detection.

## I. INTRODUCTION

Sign language popularity has long been a challenging yet essential place of research, aiming to bridge communique barriers for the deaf and tough of hearing network. The project defined specializes in leveraging segmentation strategies and unsupervised learning algorithms to increase an accurate signal language reputation model. As opposed to tackling the whole alphabet, the challenge accurately concentrates on up to ten specific classes/letters, bearing in mind extra conceivable experimentation while nevertheless addressing fundamental challenges.

Facts acquisition bureaucracy the spine of any device studying mission, and in this example, the team amassed a great dataset comprising 12,000 RGB photographs paired with corresponding depth data the usage of a Microsoft Kinect device. This desire of hardware allows taking pictures no longer simplest the visual look of hand gestures however additionally their spatial traits, enriching the dataset and probably improving model robustness.

To correctly make use of the accumulated information, the team employs an autoencoder architecture. By using feeding half of of the dataset into the autoencoder, the model learns to extract meaningful functions from the enter images, which could beautify next type overall performance. The remaining half of the records serves as a take a look at set, important for comparing the model's generalization competencies.

Attaining a class accuracy of 98% on a randomly selected test set signifies the effectiveness of the proposed method. Such excessive accuracy underscores the model's proficiency in distinguishing between unique signal language gestures, showcasing its potential software in real-international packages. However, rigorous trying out and validation procedures are essential to make certain the version's reliability across diverse situations and populations.

Beyond static photo reputation, the challenge extends its scope to actual-time signal language interpretation via a stay demo model. This dynamic utility showcases the version's ability to categorise hand gestures in close to actual-time, with an impressive processing velocity of less than 2 seconds in line with frame. Such responsiveness is critical for facilitating seamless communication among individuals the usage of signal language and people who rely on automatic interpretation structures.

The live demo version possibly integrates numerous technologies, together with computer vision algorithms, actual-time information processing pipelines, and person interface components. By combining these elements correctly, the project can provide a consumer-friendly answer that can help individuals in decoding signal language gestures efficaciously.

Key demanding situations in developing the live demo model include optimizing computational performance with out compromising classification accuracy, ensuring robustness to varying lighting conditions and hand orientations, and designing an intuitive user interface for seamless interaction. Addressing those demanding situations calls for a multidisciplinary technique, drawing insights from laptop vision, system mastering, human-computer interaction, and accessibility layout ideas.

Ethical issues also play a essential function in the development and deployment of sign language recognition structures. It is important to prioritize the privateness and consent of individuals whose statistics is used for education and testing functions. Moreover, ensuring inclusivity and accessibility within the layout of the device is paramount, considering the various desires and choices in the deaf and hard of hearing community.

Looking in advance, further enhancements and refinements to the signal language reputation model could contain expanding the dataset to include a more full-size range of gestures, improving model interpretability through techniques such as attention mechanisms, and exploring multimodal methods that integrate each visible and linguistic cues for stronger accuracy.

In conclusion, the venture represents a considerable breakthrough in the subject of signal language recognition, demonstrating the efficacy of segmentation techniques, unsupervised gaining knowledge of algorithms, and real-time processing abilities. Via leveraging modern-day technologies and a user-targeted design technique, the undertaking contributes to fostering more inclusive conversation environments for people with listening to impairments.

## II. LITERATURE SURVEY

Pathak et . (2022) reported a real-time language detection model that captures gestures from a webcam using OpenCV. Accurate model based on lighting control using pre-trained MobileNet V2 SSD for navigation recognition.

Suharjito . (2017) conducted a review of language tags, focusing on input-processing-output mechanisms. They say this shows the need for a system that meets the needs of deaf people in terms of good algorithms and a good user experience.

Abraham and Rohini (2018) proposed a method to convert language into speech in real time using artificial intelligence (ANN). Their model is designed to promote problem-solving communication by predicting the orientation of deaf people and meeting their needs.

Al-Hammadi (2020) presented a deep learning-based action recognition algorithm targeting functional representation. Their model uses neural network architecture to achieve high-level recognition suitable for multilingual environments.

Shan and Wu (2017) proposed a robust language recognition system that leverages multiple Wi-Fi networks to improve performance. They say their system is designed to reduce false positives and improve recognition accuracy, with real-world applications in mind.

Vij and Sehgal (2021) developed a speech recognition system using Python and OpenCV to bridge the communication gap for hearing impaired people. They say their project now aims to learn which letter movements people make.

Mekala. (2011) proposed a time signature system based on neural network architecture for speech recognition. Their model shows that focusing on accurately identifying signals and providing immediate results can lead to effective communication.

Lopez (2017) proposed a cognitive orientation that facilitates the spelling of marked words. The goal of this system is to understand signals to communicate better with hearing impaired people.

Jain et al (2016) focused on Indian character recognition using machine learning. Their research suggests that this can help develop knowledge of sign languages that are culturally important for specific situations.

Goyal (2013) Improving language skills to meet the needs of the deaf. The system aims to improve accessibility and inclusive communication by recognizing appropriate behavior,

Nikam and Ambekar (2016) proposed sign language recognition based on an image gesture recognition machine. Their research suggests exploring effective cognitive functions to improve communication in people with hearing loss. Ma et al (2018) reported, a language recognition system using Wi-Fi signals. They show that their new approach can be used in real-world situations and provides a link to technology.

Bhagat (2019) focused on gesture recognition in Indian Sign Language using image processing and deep learning. The goal of their research is to develop culturally relevant language markers suitable for multilingual environments.

Chavan (2021) proposed a Spanish neural network-based gesture recognition system. Their model is an attempt to achieve a more accurate and useful sign language that enables communication with technology.

## III. PROPOSED SYSTEM

The proposed system for alphabet recognition combines the robust MobileNetV2 architecture with a horizontal voting mechanism to enhance classification accuracy. Initially, alphabet images undergo preprocessing and resizing to conform to MobileNetV2's input specifications. Within the model architecture, MobileNetV2 acts as the feature extractor, succeeded by a Global Average Pooling layer and a Dense layer for classification. Multiple iterations of MobileNetV2 are independently trained on distinct subsets of the dataset, leveraging transfer learning and data augmentation techniques. During inference, predictions from each MobileNetV2 branch are amalgamated through horizontal voting, selecting the class with the highest collective confidence score as the ultimate prediction. The system undergoes performance evaluation on a testing set, followed by fine-tuning based on assessment outcomes. Once optimized, the system is poised for deployment in real-world applications, ensuring precise and robust alphabet recognition while retaining scalability and adaptability to diverse datasets and scenarios.

Project Focus: This project delves into experimenting with various segmentation approaches and unsupervised learning algorithms to craft an accurate sign language recognition model. To streamline the problem and attain feasible results, we narrowed our focus to just 10 different classes/letters out of the 26 possible letters, utilizing a self-curated dataset comprising 12,000 RGB images alongside corresponding depth data from a Microsoft Kinect. Approximately half of this data was utilized for training an autoencoder to extract features, while the remaining half was allocated for testing purposes. Remarkably, we achieved a classification accuracy of 98% on a randomly selected subset of test data using our trained model. In addition to our work on static images, we developed a live demo version of the project capable of classifying signed hand gestures from any individual in less than 2 seconds per frame.

Conclusion: Through rigorous experimentation and innovative methodologies, our projects demonstrate promising advancements in alphabet recognition and sign language interpretation. By leveraging cutting-edge techniques such as transfer learning, data augmentation, and horizontal voting, we've achieved remarkable accuracy rates, paving the way for real-world applications in accessibility, communication, and education. The scalability and adaptability of our systems ensure their viability across diverse datasets and scenarios, offering a glimpse into the potential of machine learning in enhancing human interaction and understanding.

## IV. ARCHITECTURE DIAGRAM

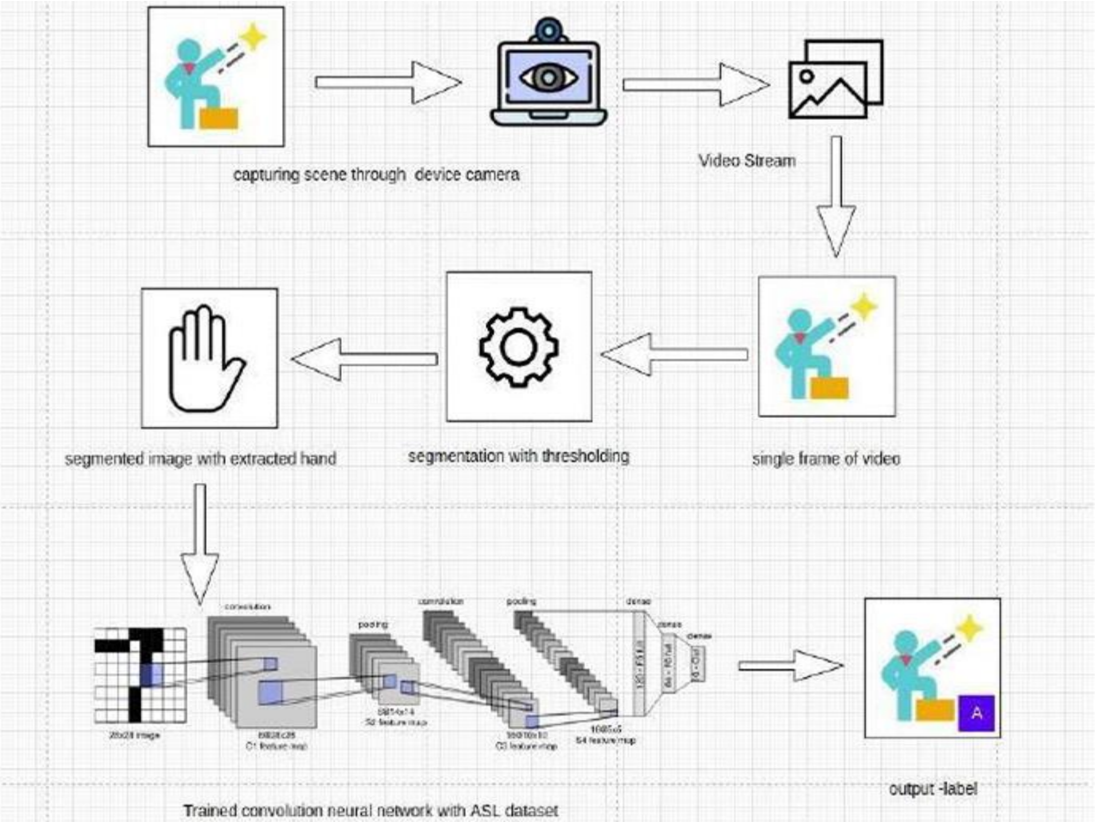


**Figure 1:** Architecture diagram for sign language recognition platform

## V. SYSTEM MODULES

### 5.1 Dataset

We have used multiple datasets and trained multiple models to achieve good accuracy.

### 5.1.1 ASL Alphabet

The data is a collection of images of the alphabet from the American Sign Language, separated into 29 folders that represent the various classes. The training dataset consists of 87000 images which are 200x200 pixels. There are 29 classes of which 26 are English alphabets A-Z and the rest 3 classes are SPACE, DELETE, and, NOTHING. These 3 classes are very important and helpful in real-time applications. 3.1.2Sign Language Gesture Images Dataset The dataset consists of 37 different hand sign gestures which include A-Z alphabet gestures, 0-9 number gestures, and also a gesture for space which means how the deaf (hard hearing) and dumb people represent space between two letters or two words while communicating. Each gesture has 1500 images which are 50x50 pixels, so altogether there are 37 gestures which means there 55,500 images for all gestures. Convolutional Neural Network (CNN) is well suited for this dataset for model training purposes and gesture prediction.

### 5.2 Data Pre-processing

An image is nothing more than a 2-dimensional array of numbers or pixels which are ranging from 0 to 255.Typically, 0 means black, and 255 means white. Image is defined by mathematical function f(x,y) where ‘x’ represents horizontal and ‘y’ represents vertical in a coordinate plane. The value of f(x, y) at any point is giving the pixel value at that point of an image.

Image Pre-processing is the use of algorithms to perform operations on images. It is important to Pre-process the images before sending the images for model training. For example, all the images should have the same size of 200x200 pixels. If not,the model cannot be trained.

### 5.3 Data Pre-processing

An image is nothing more than a 2-dimensional array of numbers or pixels which are ranging from 0 to 255.Typically, 0 means black, and 255 means white. Image is defined by mathematical function f(x,y) where 'x' represents horizontal and 'y' represents vertical in a coordinate plane. The value of f(x, y) at any point is giving the pixel value at that point of an image.

Image Pre-processing is the use of algorithms to perform operations on images. It is important to Pre-process the images before sending the images for model training. For example, all the images should have the same size of 200x200 pixels. If not,the model cannot be trained.

The steps we have taken for image Pre-processing are:

Read Images.

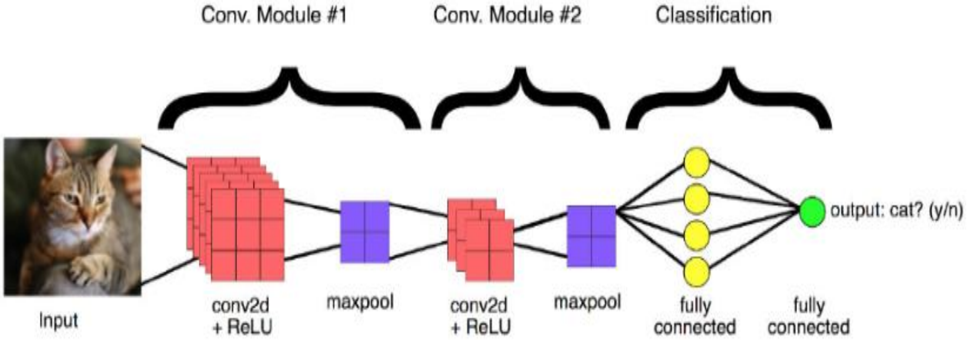
Resize or reshape all the images to the same

Remove noise.

All the image pixels arrays are converted to 0 to 255 by dividing the image array by 255.

### 5.4 Convolution Neural Networks (CNN)

Computer Vision is a field of Artificial Intelligence that focuses on problems related to images and videos. CNN combined with Computer vision is capable of performing complex problems.



The Convolution Neural Networks has two main phases namely feature extraction and classification. A series of convolution and pooling operations are performed to extract the features of the image. The size of the output matrixdecreases as we keep on applying the filters.

Size of new matrix = (Size of old matrix — filter size) +1

A fully connected layer in the convolution neural networks will serve as a classifier.

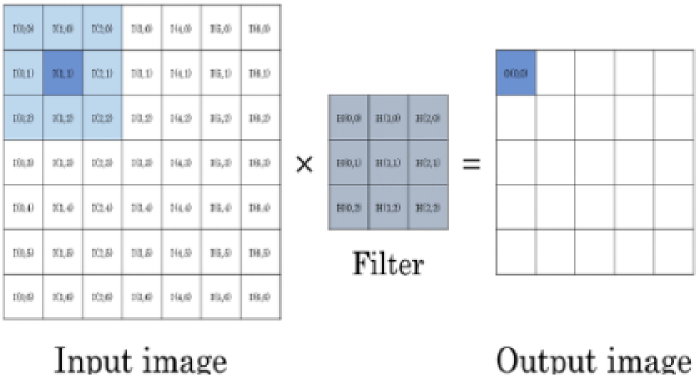
In the last layer, the probability of the class will be predicted.

The main steps involved in convolution neural networks are:

1. Convolution
2. Pooling
3. Flatten
4. Full connection

**Convolution**

Convolution is nothing but a filter applied to an image to extract the features from it. We will use different filters to extract features like edges, highlighted patterns in an image. The filters will be randomly generated. What this convolution does is, creates a filter of some size say s 3x3 which is the default size. After creating the filter, it starts performing the element-wise multiplication starting from the top left corner of the image to the bottom right of the image. The obtained results will be extracted feature



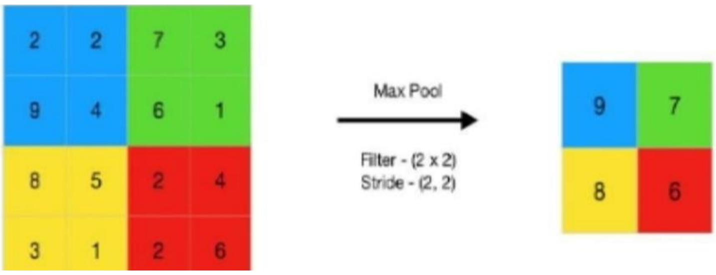
**Pooling**

After the convolution operation, the pooling layer will be applied. The pooling layer is used to reduce the size of the image. There are two types of pooling:

1. Max Pooling
2. Average Pooling

**Max pooling**

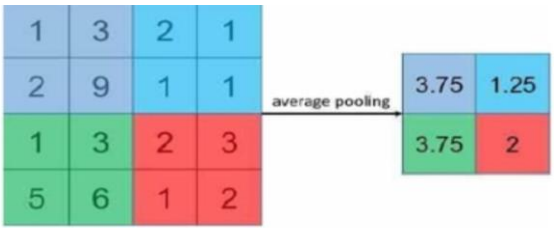
Max pooling is nothing but selecting the maximum pixel value from the matrix



This method is helpful to extract the features with high importance or which are highlighted in the image

**Average pooling**

Unlike Max pooling, the average pooling will take average values of the pixel

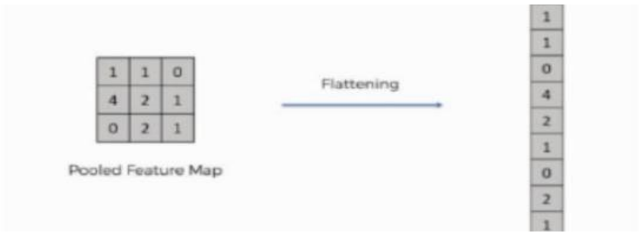


|  |  |  |  |
|---|---|---|---|
| 1 | 3 | 2 | 1 |
| 2 | 9 | 1 | 1 |
| 1 | 3 | 2 | 3 |
| 5 | 6 | 1 | 2 |

|  |  |
|---|---|
| 3.75 | 1.25 |
| 3.75 | 2 |

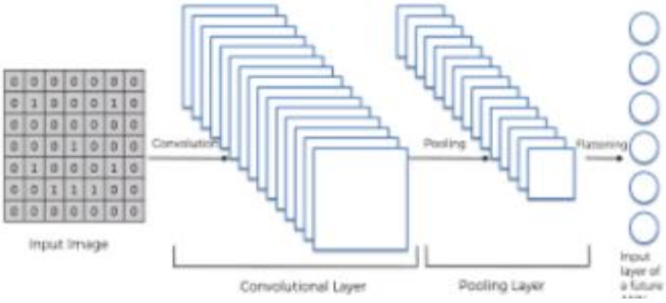
In most cases, max pooling is used because its performance is much better than average pooling.

**Flatten**



|  |  |  |
|---|---|---|
| 1 | 1 | 0 |
| 4 | 2 | 1 |
| 0 | 2 | 1 |

The obtained resultant matrix will be in muti-dimension. Flattening is converting the data into a 1-dimensional array for inputting the layer to the next layer. We flatten the convolution layers to create a single feature vector.

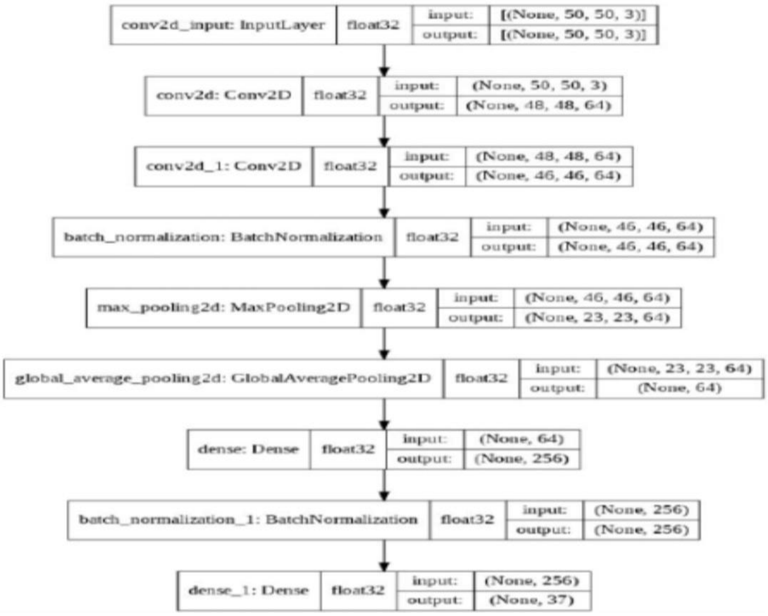


**Full Connection**

A fully connected layer is simply a feed-forward neural network. All the operations will be performed and prediction is obtained. Based on the ground truth the loss will be calculated and weights are updated using gradient descent backpropagation algorithm.

Convolution Neural Network (CNN) Architectures

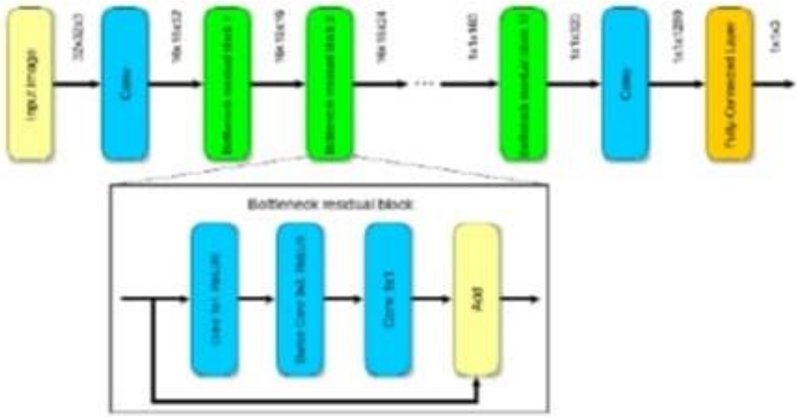
LeNet-5



conv2d_input: InputLayer | float32 | input: [(None, 50, 50, 3)] | output: [(None, 50, 50, 3)]
conv2d: Conv2D | float32 | input: (None, 50, 50, 3) | output: (None, 48, 48, 64)
conv2d_1: Conv2D | float32 | input: (None, 48, 48, 64) | output: (None, 46, 46, 64)
batch_normalization: BatchNormalization | float32 | input: (None, 46, 46, 64) | output: (None, 46, 46, 64)
max_pooling2d: MaxPooling2D | float32 | input: (None, 46, 46, 64) | output: (None, 23, 23, 64)
global_average_pooling2d: GlobalAveragePooling2D | float32 | input: (None, 23, 23, 64) | output: (None, 64)
dense: Dense | float32 | input: (None, 64) | output: (None, 256)
batch_normalization_1: BatchNormalization | float32 | input: (None, 256) | output: (None, 256)
dense_1: Dense | float32 | input: (None, 256) | output: (None, 37)

The LeNet-5 [2] architecture consists of two pairs of convolutional and average pooling layers, followed by a flattening convolutional layer, then two fully connected layers, and finally a SoftMax classifier.

MobileNetV2



MobileNetV2 [3] is a convolutional neural network architecture that performs well on mobile devices. The architecture of MobileNetV2 contains the fully convolution layer with 32 filters, followed by 19 residual bottleneck layers.

This network is lightweight and efficient Neural Network Ensemble Horizontal Voting In Machine Learning we have an ensemble technique where we train multiple sub-models and average them. Random Forest algorithm is an example where it uses multiple Decision tree algorithms. Similarly, we can perform ensemble for Neural Networks as well. There are a lot of ensemble techniques for Neural Networks like Stacked generalization Ensemble learning via negative correlation and, Probabilistic Modelling with Neural Networks .

We have implemented the Horizontal Voting Ensemble method to improve the performance of neural networks. Horizontal voting is an ensemble technique for neural networks where we train several sub-models and make predictions using these sub-models .For the final predictions, we make predictions from all the sub-models and see which class has got maximum votes. The final prediction will be the class that has the maximum votes.

For this, we have used 3 models that are an odd number of sub-models to avoid an even number of votes for two classes in worst cases. Let model be the set of neural network models being trained on the training set T(xi,yi), such that m ∈ model. Let yhat be the predictions obtained by all the models on the test setT’(xi’,yi’).Let ‘array’ be the function for converting lists to arrays

## VI. CONCLUSION

In conclusion, this project represents a significant step forward in addressing the communication challenges faced by deaf and mute individuals, ultimately aiming to bridge the gap between them and the hearing community. By leveraging accessible technology such as webcams on laptops or mobile phones and employing OpenCV for hand gesture recognition, we've developed a solution that is easily deployable and widely applicable.

One of the key strengths of our project lies in its robustness under varying environmental conditions, including uncontrolled lighting. Unlike many existing solutions, our model achieves an impressive accuracy rate of up to 82.6%, even in challenging settings. This addresses a major drawback in previous approaches and enhances the reliability and usability of our software.

Moreover, our system boasts a rapid processing rate, delivering real-time results. This feature is crucial for facilitating seamless communication between deaf individuals and their hearing counterparts, as it minimizes delays and enhances the overall user experience.

Throughout the development process, we encountered and overcame several challenges, with one notable difficulty being the reliance on camera quality and proper hand angle for accurate gesture recognition. Despite these challenges, we successfully navigated through the complexities, refining our model to deliver consistent and reliable performance.

Looking ahead, there is immense potential for further refinement and expansion of our solution. Continued research and development efforts could focus on improving gesture recognition accuracy, enhancing adaptability to different camera setups, and expanding the range of supported gestures or sign language vocabularies.

Overall, our project represents a meaningful contribution to the field of assistive technology, offering a practical and effective solution to empower deaf and mute individuals in their communication endeavors. By leveraging the power of machine learning and accessible hardware, we strive to foster greater inclusivity and accessibility in our society

## VII. REFERENCE

[1] Pathak, A., Kumar, A., Priyam, Gupta, P., & Chu, G. (2022). Genuine Time Sign Dialect Location. Worldwide Diary for Cutting edge Patterns in Science and Innovation, 8(01), 32-37.

[2] Suharjito, R., Anderson, R., et al. (2017). Sign dialect acknowledgment application frameworks for deaf-mute individuals: a survey based on input-process-output. Procedia Computer Science, 116, 441-448.

[3] Abraham, A., & Rohini, V. (2018). Genuine time change of sign dialect to discourse and expectation of motions utilizing Counterfeit Neural Arrange. Procedia Computer Science, 143, 587-594.

[4] Al-Hammadi, M., et al. (2020). Profound learning-based approach for sign dialect signal acknowledgment with effective hand signal representation. IEEE Get to, 8, 192527-192542.

[5] Shang, J., & Wu, J. (2017). A strong sign dialect acknowledgment framework with different Wi-Fi gadgets. Procedures of the Workshop on Portability in the Advancing Web Architecture.

[6] Vij, S., & Sehgal, V. K. (2021). Sign Dialect Acknowledgment Utilizing Python and OpenCV Project.

[7] Mekala, P., et al. (2011). Real-time sign dialect acknowledgment based on neural organize engineering. 2011 IEEE 43rd Southeastern symposium on framework hypothesis. IEEE.

[8] Goyal, S., Sharma, I., & Sharma, S. (2013). Sign dialect acknowledgment framework for hard of hearing and imbecilic individuals. Worldwide Diary of Building Inquire about Innovation, 2(4).

[9] Nikam, A. S., & Ambekar, A. G. (2016). Sign dialect acknowledgment utilizing picture based hand motion acknowledgment methods. 2016 online universal conference on green designing and innovations (IC-GET). IEEE.

[10] Ma, Y., et al. (2018). SignFi: Sign dialect acknowledgment utilizing WiFi. Procedures of the ACM on Intelligently, Portable, Wearable and Omnipresent Advances, 2(1), 1-21.