International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 24 Issue 05, MAY, 2024

INDIAN SIGN LANGUAGE RECOGNITION USING OPEN CV

D.Varaprasada padmaja[1], V.Meghana[2], T.satvika[3], P.N.rohitha[4], CH.V.Durga[5]

[1]Assisant Professor, Department of CSE-Artificial Intelligence and Machine Learning , S.R.K Institute of Technology, NTR, Andhra Pradesh, India, meghana04veluri@gmail.com
[2,3,4,5]*student, Department of CSE-Artificial Intelligence and Machine Learning, S.R.K Institute of Technology, NTR, Andhra Pradesh, India*

Abstract:

This project proposes the development of an Indian Sign Language Recognition System using Open-Source Computer Vision Library. The system aims to bridge the communication gap between individuals with hearing impairments who use ISL and those who do not understand the language. Leveraging computer vision techniques, the system will detect and interpret ISL gestures and movements in real-time, enabling the translation of sign language into text or speech. The project seeks to address the lack of accessible tools for learning and using ISL, thereby promoting inclusivity and enhancing communication accessibility for the hearing-impaired community in India. Sign languages are defined as an organized collection of hand gestures having specific meanings which are employed from the hearing-impaired people to communicate in everyday life.

Communication is an important aspect when it comes to share or express information, feelings, and it brings people closer to each other with better understanding. Sign language, a full-fledged natural language that conveys meaning through gestures, is the primary chief of communication among Deaf and Dumb people. A gesture is a pattern which may be static, dynamic or both, and is a form of nonverbal communication in which bodily motions convey information. Sign language translation is a task forautomatically translating sign languages into written languages which is already existed.Now we are going to implement a system which is used to convert the text which is produced sign language translator into speech. In this project we are going to implementa deep learning algorithms-based system such as CNN for translation of text (i.e., which is extracted from sign language) into speech. CNN are to capture intricate hand movements and to learn the temporal relationships between the hand gestures respectively. Later the translated text is then converted to speech using a Text- To-Speech (TTS) API. This allows the system to provide a complete communication solution for deaf and mute individuals.

1. Introduction:

In order to communicate knowledge in our everyday lives, contact between various communities is essential and crucial. Effective communication is an essential life skill, but individuals with speech and hearing impairments find it challenging to share their thoughts with others. There are several ways that two individuals can communicate with each other. Not everyone is able to decipher sign language while conversing with members of the deaf and dumb communities. It is challenging to communicate without the assistance of an interpreter or other resources. In order to facilitate barrier-free communication and make sign language understandable to others, we must translate it.

A system for recognising sign language is one of the efficient solutions to this problem. Sign language uses a variety of hand movements to convey important information. There are regional variations in sign language around the globe. Around the world, 135 sign languages are commonly used for communication. Every sign language is distinct from the others, much as American Sign Language differs from Indian sign language in India. We standardised to focus on Indian Sign Language motions because we thought it would be easier to interpret. In order to enable barrier-free communication between them and others, we must translate Indian sign language.

The identification of sign language remains a difficult topic and has been the focus of much study in recent years. Using hand gloves for computer-human interaction is one way to recognise hand gestures. However, the complexity of this approach lies in the fact that the user must wear gloves and carry a heavy bundle of cables to connect the gadget to the computer. Therefore, we suggested working on sign recognition with just our hands—that is, without the need for any extra wearable hardware—in order to remove this complexity and make user contact with computers simple and natural. Processes for recognising sign language mostly rely heavily on human translation services. Involving human knowledge in translation is likewise highly costly and complex. Now, let us present our automated sign language proposal.

Several components, including object tracking, skin segmentation, feature extraction, and recognition, are present in all sign language recognition systems. The primary functions of the first two modules are to find and extract hands from video frames, while the subsequent modules are utilised for gesture identification, feature extraction, and classification. Since there are many variables in the picture space for an image-based gesture recognition system, it is important to extract the image's key properties. Our project's main goal is to create a model that can identify hand movements based on finger spelling and combine each motion to produce a whole word.

The deaf community makes great use of language translators to transform and form their thinking. There is an immediate need for a system that can translate and recognise sign language. The absence of an effective gesture recognition system created especially for people with disabilities inspires us as a group to make significant progress in this area. The goal of the proposed effort is to translate these sign motions into speech that the average person can understand. CNN architecture develops the full model pipeline for the categorization of 26 alphabets plus one additional alphabet for the null letter.

Our approach has a great degree of efficiency when it comes to real-time gesture prediction from sign language. These anticipated alphabets are transformed into words, which then create sentences. As a result, this approach may be applied in real-time to help close the communication gap that exists between the Deaf and Dumb and the general public.

2. Existing System:

Sign language translation is still in its infancy compared to other sign languages, it is one of the most difficult subjects to study. The study has demonstrated the use of machine learning models for sign language categorization.

Thus, standard data sets with variances and disturbances are quite rare. It causes feature occlusion, which is a significant hindrance to the lack of advancement in this area. By offering a data set of sign language translation, the current effort seeks to enhance the study in this area. We developed a sign language database for both alphabetic and numeric characters.

The characteristics will then be retrieved from the gathered segmented data using the Bag of Words model and picture pre-processing.

3.Literature Survey

1.Shravani K,etal IOSR Journal of Computer Engineering (IOSR-JCE), 22(3), (2020), pp. 14-19

The journal provides insights into the nature and significance of gestures, particularly within the context of sign language. It defines gesture as a form of non-verbal communication characterized by bodily motions that convey information, which can be either static (unchanging) or dynamic (changing over time). Sign language, as described, encompasses visual gestures and signs used by individuals who are deaf or mute to communicate. It emphasizes that sign language is a structured code, where each sign carries a specific meaning assigned to it. These signs go beyond representing just alphabets or numbers; they also convey common expressions, greetings, and full sentences, allowing for rich and nuanced communication.

Furthermore, the journal highlights the distinction between different sign languages, such as Indian Sign Language (ISL) and American Sign Language (ASL). It notes that ISL utilizes gestures involving both hands and is considered more complex compared to ASL. This complexity may arise from the intricacies of representing meaning through simultaneous movements of both hands. However, the journal also points out that the complexity of ISL may have led to relatively less research and development in this field compared to ASL. Overall, the journal underscores the importance of gestures and sign language as fundamental modes of communication for individuals who are deaf or mute. It highlights the structured nature of sign language and acknowledges the complexities associated with representing meaning through gestures, particularly in ISL. Additionally, it suggests a need for more research and development effortsto further understand and advance the field of sign language, particularly in the context of ISL

2. Babita Sonare, Aditya Padgal, Yash Gaikwad, Aniket Patil Department of Information Technology" Pimpri Chinchwad College of Engineering, May 2021.

It highlights the importance of developing an interactive, real-time video-based sign language translation system, particularly tailored for individuals who are deafor mute and face challenges in communicating with others. Such a system, poweredby efficient machine learning algorithms, holds significant potential to bridge the communication gap between individuals with hearing and speech impairments and those who can hear and speak. Central to the development of such a system is the recognition of gestures and human activity, both of which play crucial roles in detecting and interpreting sign language as well as understanding the behavior of individuals. Gesture recognition involves the identification and analysis of hand movements, facial expressions, andbody postures, which are fundamental components of sign language communication. Human activity recognition, on the other hand, encompasses the broader context of human actions and interactions, providing insights into the intentions and behaviors of individuals. These domains, gesture recognition, and human activity recognition, are rapidly advancing areas of research and development. They not only contribute to the creation of sign language translation systems but also find applications in various other fields, including automation in households and industries. The integration of efficient machine learning algorithms into these systems enables higher levels of automation and efficiency, facilitating seamless communication and interaction between individuals with hearing and speech impairments and their counterparts in both personal and professional settings.

The development of an interactive, real-time video-based sign language translationsystem powered by efficient machine learning algorithms holds great promise for improving communication accessibility for individuals who are deaf or mute. By leveraging advancements in gesture and human activity recognition, these systemspave the way for greater inclusivity and automation in diverse contexts, benefitingboth individuals and society

3. K. Amrutha, P. Prabu 2021 International Conference on Innovative Trends in Information Technology (ICITIIT).

The development of the model centres on vision-based isolated hand gesture detection and recognition, aiming to provide a solution for individuals with speechand hearing impairments to effectively communicate through sign language. By segmenting sign language into region-wise divisions, the model offers a straightforward method for users to convey information, enhancing accessibility and understanding. This approach is particularly valuable considering that a significant portion of society does not comprehend sign language, leaving speech and hearing-impaired individuals reliant on human translators for communication. However, the availability and affordability of human interpreters may be limited, presenting challenges in ensuring consistent and accessible communication. To address these challenges, an automated translator system emerges as a viable solution, capable of interpreting sign language and converting it into a comprehensible format. Such a system would significantly reduce the communication gap that exists among individuals in society, empowering speech, and hearing-impaired individuals to communicate more effectively with others. Byleveraging vision-based technology and machine learning algorithms, thetranslator system can accurately detect and recognize hand gestures, facilitating seamless communication without the need for human intervention. Overall, the development of an automated sign language translator represents a significant step towards fostering inclusivity and accessibility in society. By providing a reliable and affordable means of communication for speech andhearing-impaired individuals, the translator system contributes to breaking down barriers and promoting greater understanding and connection among people from diverse backgrounds.

4. Aashir Hafeez, Suryansh Singh, Ujjwal Singh, Priyanshu Agarwal, Anant KumarJayswal.

Amity School of Engineering and Technology Amity University, Noida Uttar Pradesh, India.The journal highlights the prevalent use of sign language among the majority of deaf individuals as their primary mode of communication. It underscores the challenge faced by those who do not understand sign language in effectively interacting with individuals who rely on it for communication. In response to thischallenge, researchers have developed a device known as a sign language recognition system (SLR). The study described in the journal focuses on comparing various machinelearning techniques using a dataset specifically designed for American Sign Language (ASL). This dataset likely contains a collection of images or videos capturing different ASL gestures and corresponding labels. By leveraging this dataset, researchers can train and evaluate machine learning models to recognize ASL gestures accurately. The journal delves into the multiple stages involved in the development of an automated SLR system. These stages typically include data collection, pre-processing, feature extraction, model training, evaluation, and deployment.

Data collection involves gathering a comprehensive dataset of ASL gestures, while preprocessing involves tasks such as image or video cleaning, normalization, and segmentation. Feature extraction aims to extract relevant features from the data, such as hand shapes, movements, and orientations.

Model training involves utilizing machine learning algorithms to train models onthe extracted features, while evaluation assesses the performance of these modelsusing metrics such as accuracy, precision, recall, and F1-score. Finally, deployment involves integrating the trained model into a real-world application or device, such as a mobile app or a wearable device, to enable real-time ASL gesture recognition.

5. U & P U Patel Department of Computer Engineering, CSPIT, CHARUSAT Campus, Charotar University of Science and Technology (CHARUSAT), Changa 388421, India.

The predominant means of communication is speech; however, there are persons whose speaking or hearing abilities are impaired. Communication presents a significant barrier for persons with such disabilities. The use of deep learning methods can help to reduce communication barriers. This paper proposes a deep learning-based model that detects and recognizes the words from a person’s gestures. Deep learning models, namely, LSTM and GRU (feedback-based learning models), are used to recognize signs from isolated Indian Sign

Language (ISL) video frames. The four different sequential combinations of LSTM and GRU (as there are two layers of LSTM and two layers of GRU) were used with our own dataset, IISL2020. The proposed model, consisting of a single layer of LSTM followed by GRU, achieves around 97% accuracy over 11 different signs.

Research has been conducted on voice generation using smart gloves, which could give a voice to sign language movements. However, those who do not know sign language usually undervalue or reject persons with such an impairment because of the lack of proper communication between them. Hence, this paper proposes a system aimed at removing the communication gap and giving every individual a fair and equal chance. It involves taking a video of the person making hand gestures, processing it, and passing it to the proposed model, which predicts words one by one. The system then generates a meaningful sentence out of those words that can then be converted into the language selected by the communicator. There are more than 120 distinct sign languages, such as American sign language, Indian Sign Language, Italian sign language, etc. In the proposed system, the authors used Indian sign language to provide an example. The dataset for this system was customized; there are 11 static signs and 630 samples in total. The signs that were included are commonly used words, such as hello, goodbye, good morning, etc. The system was designed so that it uses the natural gesture input to create a sign language and then passes this to the system for further pre-processing and processing tasks to predict the exact word that the gesture expressed.

6.Department of Computer Science and Technology, Manav Rachna University, Faridabad 121010, Haryana, India by **Harsh Kumar Vashisth Tuhin Tarafder Rehan Aziz Mamta Arora** and **Alpana**

Rosalina et al. have taken about 3900 raw image files to achieve the same with over 39 alphabets, numbers, and punctuation marks, in accordance with SIBI (Sistem Isyarat Bahasa Indonesia), and the accuracy turned out to be 90%. Computer vision were used to capture the image and then extract essential data from it. ANN (Artificial Neural Network) was then used to classify the images, and at last, speech recognition was used to translate the input speech in the form of NATO phonetic language and then translate it to sign language. Image processing is conducted by morphological operators, mainly erosion and dilation. The image is to be segmented. The whole process consists of four stages: image capture (static photos in RGB color space), image processing (separate background from hand, HSV range to separate the same, blurred, then dilated), feature extraction (finding contours that are basically edges of the same color range), and classification (ANN will take B/W (Black and White) images). Lastly, they translated the given letters into speech, which can be conducted using the NATO alphabets.

B. Hangün et al.compare the performance of functions related to image processing as implemented in the OpenCV library. Image processing requires more computational power than regular use because of mathematical operations such as matrix inversion, transposition of a matrix, matrix convolution, Fourier transform, etc. Images are taken as matrices, and as the image resolution increases, so does the matrix order. CPUs and GPUs work on different principles; CPUs operate on series processing (one task at a time) and GPUs on parallel processing (multiple tasks at the same time). Although this does not completely hold true today, it is still practically true.

4. Proposed System

When it comes to sharing or expressing thoughts and ideas, communication is crucial since it improves understanding between individuals. The main form of communication between Deaf and Dumb people is sign language, a full-fledged natural language that uses movements to communicate meaning. In this project, we'll put in place a text translation system that uses deep learning techniques, such CNN deep learning model.The purpose of CNN respectively, is to understand the temporal correlations between the hand motions and to capture complex hand movements. This makes it possible for the system to offer deaf and mute people a comprehensive communication option.

I) Dataset:

Data Set:

Using 17113 pictures, the system taught CNNs to classify integers, alphabets, and other commonly used terms. With our approach, 96% of the alphabet's 27 letters are correctly identified. The outcome also demonstrates that the accuracy of the system increases as the number of images—which may include pre-processed images—in the dataset increases.

II Data Preprocessing:

Background Subtraction:

If applicable, this technique removes background elements from the image, isolating the primary object of interest.

Grayscale Conversion:

Images are converted to grayscale to simplify processing, as color often doesn't contribute significant information for classification tasks.

Canny Edge Detector:

By emphasising the edges and contours of the items in the image, this technique produces a simpler representation that is easy for the model to understand.

Features:

Our approach has a great degree of efficiency when it comes to real-time gesture prediction from sign language. These anticipated alphabets are transformed into words, which then create sentences. Because we employed a small CNN-based architecture, the model is computationally efficient and effective, which facilitates its deployment on embedded platforms like as Raspberry Pi and Google Coral. As a result, this approach may be applied in real-time to help close the communication gap that exists between the Deaf and Dumb and the general public.

Data Acquisition:

Despite our best efforts, we were unable to collect our dataset. As a result, we had to apply our pre-processing approach straight to the pre-existing dataset.

Pre-processing:

The amount of data needed for the model to function well during training is rather high. Therefore, if the number of pictures in our dataset for our network is restricted, we have often supplemented our photos in order to enlarge the data set. Our dataset has only undergone small changes, such as flips, shifts, and rotations. Additionally, data augmentation might lessen the likelihood that a model would overfit. In order to handle every image equally, we have shrunk and rescaled our photographs here.

Feature Extraction:

As a pre-processing method, the Gaussian filter is used to smooth out the picture and remove any irrelevant noise. To eliminate spurious edges, intensity is examined and non-maximum suppression is used.

Double thresholding is used to take into account just the pictures' strongest edges in order to improve the pre-processed image data. Finally, all of the weak edges are eliminated, leaving just the strong edges to be taken into consideration for the next stages.

III Recognition:

In order to do this, a trained model was loaded onto a laptop using TensorFlow as the backend. Real-time hand-shaped video frames were then collected using OpenCV. As a result, the model accurately recognises and predicts the input hand motions.

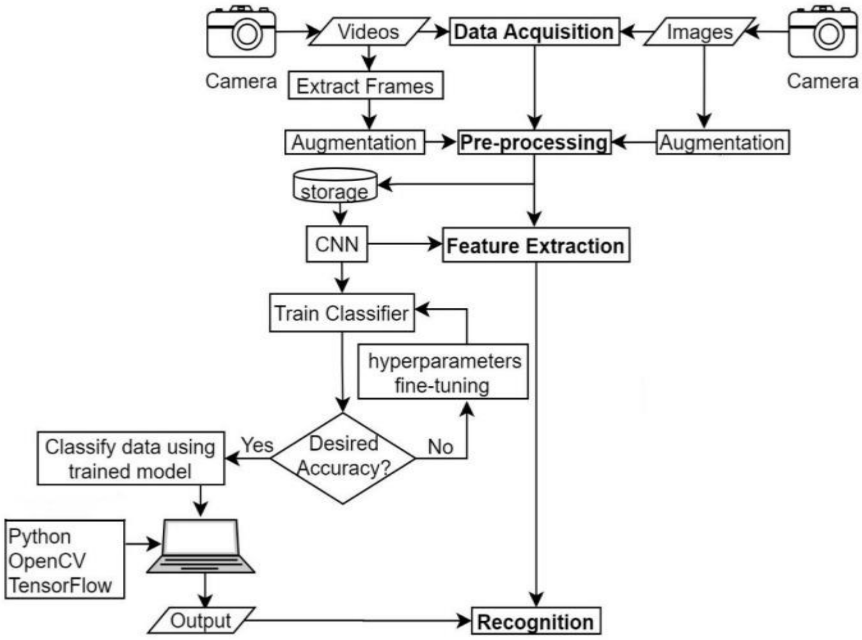

Fig 1: System Architecture

4.1 Methodology:

1. Data Collection:

Gather a dataset of images or videos containing various ISL gestures. Ensure diversity in lighting conditions, backgrounds, and hand orientations to make your model robust.

2. Preprocessing:

Convert the collected data to grayscale to simplify processing. Apply techniques like noise reduction, image smoothing, and contrast enhancement to improve the quality of images.

3. Hand Detection:

Use techniques like background subtraction or Haar cascades to detect and localize hands in the input images or video frames.

4. Hand Segmentation:

Segment the detected hand region from the rest of the image using techniques like thresholding, contour detection, or skin color detection.

5. Feature Extraction:

Extract relevant features from the segmented hand region. This can include:

Hand shape descriptors (e.g., Hu moments, contours).

Hand motion features (e.g., optical flow, temporal derivatives).

Hand position and orientation.

6.Model Training:

Train a machine learning model or a deep learning model to classify the hand gestures. Common approaches include:

Convolutional Neural Networks (CNNs) trained end-to-end on raw pixel data or extracted features.

7.Model Evaluation:

Evaluate the trained model using appropriate metrics like accuracy, precision, recall, and F1-score on a separate validation dataset.

8.Integration with OpenCV:

Implement the trained model within an OpenCV framework to perform real-time gesture recognition on live video streams or recorded videos.

9.Testing and Optimization:

Test the integrated system on various real-world scenarios and optimize parameters like thresholds, model architecture, and hyper parameters to improve performance.

4.2 Algorithms Used:

Convolutional Neural Network (CNN):

Unlike regular Neural Networks, in the layers of CNN, the neurons are arranged in 3 dimensions: width, height, depth. The neurons in a layer will only be connected to a small region of the layer (window size) before it, instead of all of the neurons in a fully-connected manner. Moreover, the final output layer would have dimensions (number of classes), because by the end of the CNN architecture we will reduce the full image into a single vector of class scores.

CNN Model:

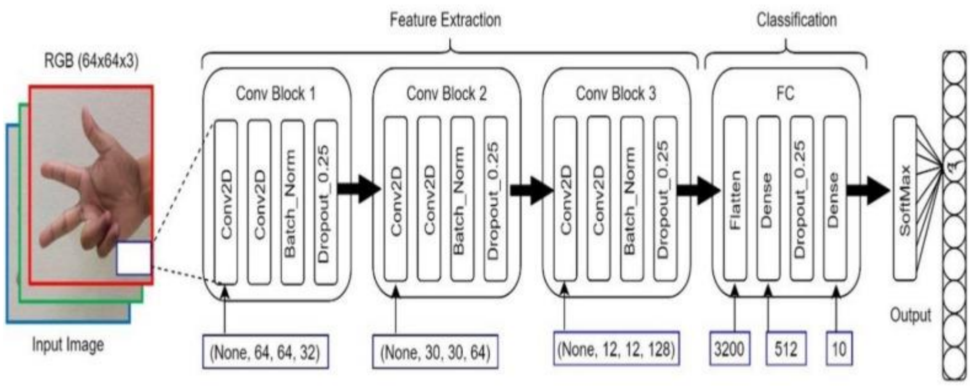

FIG 2: CNN STRUCTURE FOR HAND GESTURE

1st Convolution Layer: The input picture has resolution of 200x200 pixels. It is first processed in the first convolutional layer using 64 filter weights.

1st Pooling Layer: The pictures are down sampled using max pooling of 3x3 i.e., we keep the highest value in the 3x3 square of array. Therefore, our picture is down sampled.

2nd Convolution Layer: Now, this output of the first pooling layer is served as an input to the second convolutional layer. It is processed in the second convolutional layer using 128 filter weights(2x2 pixels each).

2nd Pooling Layer: The resulting images are down sampled again using max pool of 3x3 and is reduced to even lesser resolution of image.

3rd Convolution Layer: convolutional layer using 256 filter weights (2x2 pixels each).

3rd Pooling Layer: The resulting images are down sampled again using max pool of 3x3 and is reduced to even lesser resolution of image.

Flatten Layer:

It is used to convert the 2D pixel array into linear form in order to produce converge it into 27 class of hand signs.

Final layer:

The output of the 3rd Densely Connected Layer serves as an input for the final layer which will have the number of neurons as the number of classes we are classifying (alphabets + blank symbol).

Activation Function:

We have used ReLu (Rectified Linear Unit) in each of the layers (convolutional as well as fully connected neurons). ReLu calculates max(x,0) for each input pixel. This adds nonlinearity to the formula and helps to learn more complicated features. It helps in removing the vanishing gradient problem and speeding up the training by reducing the computation time.

At the last activation function, we used SOFTMAX function. It is used as the activation function in the output layer of neural network models that predict a multinomial probability distribution. That is, SoftMax is used as the activation function for multi-class classification problems where class membership is required on more than two class labels. •

Pooling Layer:

We apply Max pooling to the input image with a pool size of (3, 3) with ReLu activation function. This reduces the amount of parameters thus lessening the computation cost and reduces overfitting.

Dropout Layers:

The problem of overfitting, where after training, the weights of the network are so tuned to the training examples they are given that the network doesn't perform well when given new examples.

5 .OUTPUTS SCREENSHOTS :

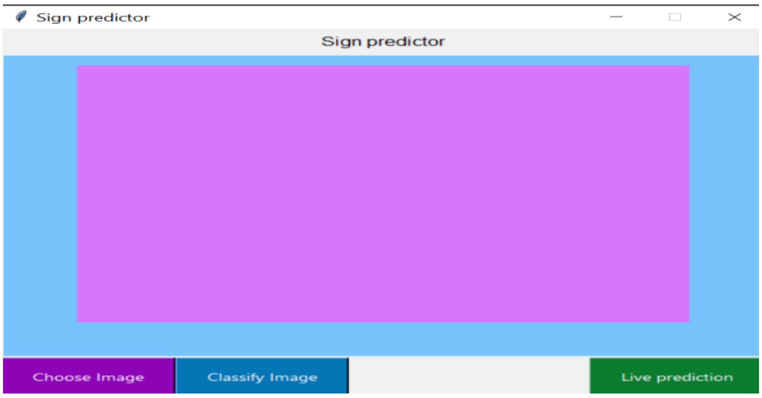

FIGURE 3 HOME PAGE

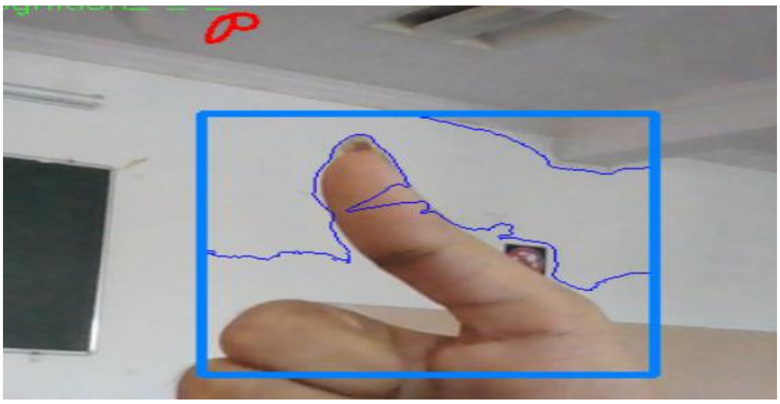

FIGURE 4 OUTPUT

FIGURE 5 OUTPUT

6. Conclusion:

Communication between deaf-mute and a normal person have always been a challenging task. The goal of our project is to reduce the barrier between them. We have made our effort by contributing to the field of Sign Language recognition. In this project, we developed a CNN-based human hand gesture recognition system. The salient feature of our system is that there is no need to build a model for every gesture using hand features such as fingertips and contours. Here in this project, we have constructed a CNN classifier which is capable of recognizing sign language gestures. The proposed system has shown satisfactory results on the transitive gestures. In this report, a functional real time vision-based sign language recognition for deaf and dumb people have been developed. We achieved final accuracy of 98.0% on our dataset. We are able to improve our prediction after implementing two layers of algorithms, we have also verified our result for the similar looking gesture which were more prone to misclassification. This way we are able to detect almost all the symbols provided that they are shown properly, there is no noise in the background and lighting is adequate.

7. Future Scope:

The future scope of the project involving Artificial Neural Networks (ANNs) is vast and promising, with numerous avenues for advancement and innovation. Potential directions include exploring more complex neural network architectures such as deep neural networks (DNNs) and recurrent neural networks (RNNs) to improveperformance in tasks like image classification, speech recognition, and natural languageprocessing. Additionally, integrating ANNs with emerging technologies like augmented reality (AR) and virtual reality (VR) could lead to novel applications in fields such as education, healthcare, and entertainment. Further enhancements could beachieved through multimodal learning approaches, real-time interaction capabilities, and a focus on accessibility and inclusivity for diverse user groups. Ethical considerations and responsible AI deployment will also play a crucial role in shaping the future development and deployment of ANN-based systems, ensuring fairness, transparency, and accountability in their implementation. Overall, the future of the project holds immense potential for addressing complex real-world challenges and making meaningful contributions to various industries and domains.

8.REFERENCES:

[1] K. M. J. R. A. &. R. I. Tiku, "Real-time Conversion of Sign Language to Text and Speech," in Tiku, K., Maloo, J., Ramesh, A., & R, I. (2020). Real-time Conversion of Sign Language to Text and Speech. 2020 Second International CSecond International Conference on Inventive Research in Computing Applications (ICIRCA), 2020.

[2] S. Y. M. M. K. S. V. S. &. S. S. Heera, "Talking Hands – An Indian Sign Language to Speech Translating Gloves," in International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2017), 2017**.**

[3] Hunter Phillips, Steven Lasch & Mahesh Maddumala , "American Sign Language Translation Using Transfer Learning".

[4]M. Rajmohan, C. Srinivasan, Orsu Ranga Babu, Subbiah Murugan, Badam Sai Kumar Reddy "Efficient Indian Sign Language Interpreter For Hearing Impaired".

[5]Mahmudul Haque, Syma Afsha, Tareque Bashar Ovi, Hussain Nyeem, "Improving Automatic Sign Language Translation with Image Binarisation and Deep Learning"

6] K.Bhanu Prathap, G.Divya Swaroop, B.Praveen Kumar, Vipin Kamble, Mayur Parate, "ISLR: Indian Sign Language Recognition" .

[7] Pavleen Kaur, Payel Ganguly, Saumya Verma, Neha Bansal, "Bridging the Communication Gap: With Real Time Sign Language Translation".

[8]Hao Zhou, Wengang Zhou, Weizhen Qi, Junfu Pu, Houqiang Li, “Improving Sign Language Translation with Monolingual Data by Sign Back-Translation”.
[9] Wanbo Li, Hang Pu, Ruijuan Wang, “Sign Language Recognition Based on Computer Vision”.
[10]Neeraj Kumar Pandey, Aakanchha Dwivedi, Mukul Sharma, Arpit Bansal, Amit Kumar Mishra, “An Improved Sign Language Translation approach using KNN in Deep Learning Environment”.
[11]. R Vijaya Prakash, Akshay R, A Ashwitha Reddy, R Harshitha, K Himansee, S.K Abdul Sattar, “Sign Language Recognition Using CNN”.
[12]. Sakshi Sharma, Sukhwinder Singh, “Vision-based sign language recognition system: A Comprehensive Review