International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP. 67-71

Two Way Smart Communication System for Deaf andDumb People

Deekshith R[1], Charan Kumar L[2], Chhanukya K.V[3], Darshan Ekbote[4], K.N Pushpalatha[5]

[1]*(Department of ECE, Dayananda Sagar College of Engineering,India)*
[2]*(Department of ECE, Dayananda Sagar College of Engineering,India)*
[3]*(Department of ECE, Dayananda Sagar College of Engineering,India)*
[4]*(Department of ECE, Dayananda Sagar College of Engineering,India)*

Abstract: *For the deaf and dumb community, sign language could be a primary form of communication. This language specifies its information, among other things, using a set of representations, which are finger signs. This system offers a novel technique for real-time applications, consisting of two parts: the first is gesture action analysis, detection, and generation of a text description and its associated sound; and the second is for a typical user who can quickly respond to them through speech. Voice analysis, text conversion, and related movements are employed as outputs. The movements are learned using a convolutional neural network model, while the speech is transformed to text and both the text and gestures are presented using Python programming.*

Keywords - *Convolution neural network, Deaf, Dumb, Gestures, Recognition, Sign Language, Text conversion*

I. Introduction:

India has around 2.5 million deaf and dumb people, accounting for 25 percent of the world's deaf and dumb population. These folks lack the basic necessities that a normal person would have. Lack of communication could be a major factor, as deaf people are unable to focus and dumb people are unable to talk. Deaf and stupid persons account for more than 5% of the population. The Deaf and Dumb primarily communicate with each other through communication. The most significant disadvantage that Deaf and Dumb people confront nowadays is conversing with those who do not understand communication. The physical impairment of hearing for deaf people and the disability of speaking for dumb people may be the cause of the declining ratio of literate and employed Deaf and Dumb people. This will result in a communication barrier between non-disabled people and Deaf and Mute people It’s very difficult for the people of two different communities with two different languages to communicate, thus communicating with one another becomes a stumbling block, and they need a translator physically, which isn't always convenient to arrange.[1]The conventional person and the hard to hear kind of person have the same dilemma. Writing, on the other hand, is connected with possibility; it is considered a sluggish and inefficient mode of communication. As a result, hiring a professional language translator is a reasonable option. For deaf, dumb, and normal individuals, we're developing a two-way smart communication system. The aim of the project is to create a sophisticated communication system that will let deaf and hard of hearing persons communicate with others who are not deaf or hard of hearing. A regular individual can also respond to their messages. We have a clever communication system that is two-way. The technique to communicate in this arrangement is for a Deaf and Dumb person to sit down with a typical person. Image processing takes a variety of inputs, such as an image or a collection of frames from a video, and produces an image or parameters that are connected to an image as output. The technique “Image processing” is used to improve image quality and extract usable information from images. This is referred to as "feature extraction." [2] These image processing techniques are used for a variety of applications, including identifying hand movements, evaluating actions, and more. Image processing techniques are utilised in these systems for communication in a community where people are deaf and dumb.

II. Literature Survey:

Ishika Dhal et al. [1] Using a Deep Convolutional Neural Network model, we were able to recognize handgestures automatically. It includes a cutting-edge deep Neural Network and Multi-layer Neural Networks for hand sign gesture identification .The labeled dataset that is fed and trained in the models. These models are utilized to match with the desired output while identifying the images on dynamic images or still images. Prashant G. Ahire et al. [2] Deaf and dumb people, as well as normal people, have a two-way communication system. The system is divided into two components. The initial step is to extract Indian Sign Language (ISL) gestures from real-time video and map them to speech that can be understood by humans. As a result, the second section will accept normal English as input and convert it to animated Sign Language motions. Frame construction from films, detecting regions of interest and mapping of visuals with a language knowledge base using a correlation based method will all be part of the video to speech processing, which will be followed by appropriate audio production utilizing the Google Text-to-Speech API. By converting audio to text using the Google Speech-to-Text API and then mapping the text to relevant animated gestures from the database, natural language is used to transfer the equivalent Indian Language gestures. Felix Zhan et al. [3] Recognition of Human Hand Gestures Convolutional Neural Networks Intelligent video systems are developed using automatic human gesture detection from camera images. From a camera image, this system presents a convolution neural network method for recognising hand motions of human task activities. The skin model and hand position and orientation calibration are used to acquire training and testing data for the CNN in order to attain excellent performance. Because light has a significant impact on skin color, we use a Gaussian Mixture model to train the skin model. The goal of hand position and orientation calibration is to convert the hand image into a neutral pose. The CNN is then trained using the calibrated pictures.

III. Methodology and Implementation:

3.1 Gesture to speech conversion:

A convolutional neural network is a deep learning neural network class that is most often applied to images and videos for its analysis. It is a kind of artificial neural network using machine learning algorithms for a unit and perceptrons for supervised problems. A Convolution neural network (CNN) is basically a technique or a machine learning model applied to images to make them understandable by machines. It can have one or more than one convolutional layer followed by the fully connected layers. All types of cognitive tasks are performed using CNNs, like Natural Language processing, image processing, etc. The concept of machine learning is not a contemporary thing.[3] The first Artificial Intelligence-based programme came intoplay with a learned version of a game in which an AI programme was built that understood natural language.

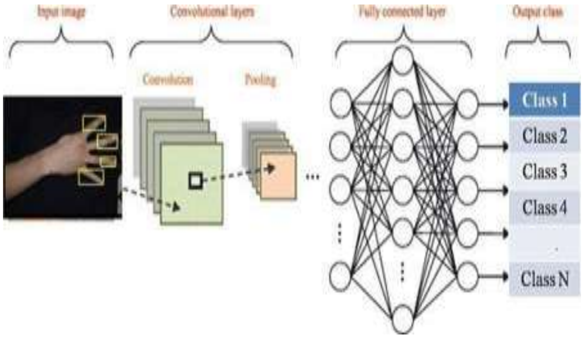


Fig 1. Working of a CNN Model

A CNN, (convolutional neural networks) is also known Artificial neural Network, that is used to analyze the videos and pictures. It is mainly used for feature extraction. It's a type of artificial neural network. Convolution neural networks use some machine learning algorithms (models) and apply these techniques to images and videos to make images comprehensible to machines. A convolution neural network has one or many convolution layers, then followed by the rest of the other layers, such as the pooling layer and the Fully connected layer. All types of processing tasks, such as video processing, image processing, and language processing, can be performed by using convolution neural networks (CNN). The machine learning concept can be widely used in various types ofapplications.

There are different types of layers in Convolutional neural network(CNN) model such as:

i. Input layer
ii. Hidden layer:
a. Convolutional Layer
b. Pooling Layer
c. Fully Connected Layer
iii. Output layer.

There are three different layers that come under the hidden layer, such as the convolution layer, pooling layer, and fully connected layer. The convolution layer is the first layer which takes the input image from hand,

and its primary goal is to detect and extract features from the image. After extracting the feature from the image, an input matrix is generated. This layer is used to find the edges of an input image. The convolutional process is a mathematical function that detects the edges and tiny patterns in an image. Then convert a single frame to a matrix of discrete values (0’s and 1’s) in which the value 0 is assigned to all areas of the image that are black and 1 is assigned to all areas of the image that are white. Integer values from 0 to 255 are used to create a matrix. All the values in the matrix represent shades in a grayscale image. But if the input image is in color format, then a three-channel image is used. Then we select a different filter with random values. Next, we define the filter matrix, which is also known as the kernel matrix, that contains 0’s and 1’s in the particular order. By using both the input matrix and the kernel matrix, we generate a new matrix with all of the edges as shown in figure 2. To find a fresh image with just the borders, filtering of the image is required. In order to find the scalar product the technique needs to be applied on the image’s pixels.

The image then passes to the training phase. As they progress through the training phase, the numerical data present in the matrix are upgraded to their convenient values. Every matrix value is multiplied by the dot product of pixels. After this, move this kernel matrix upon the single frame and look for the convoluted image matrix. Multiply the kernel matrices and the image matrices element-by-element. Attempt to use many kernels in a single picture. When numerous kernels are applied to a single image, multiple convolved matrices arise. Finally, a new convolved matrix is generated, which is shown in fig.2.

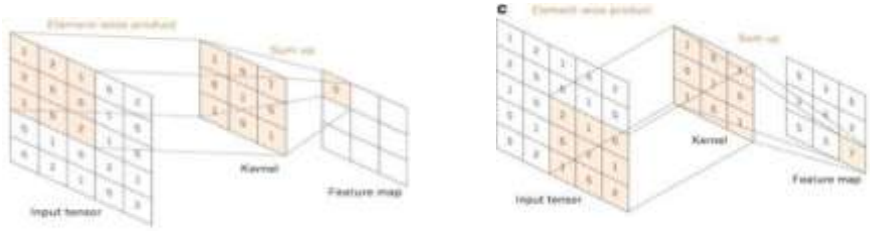


Fig 2. Extracting convolved matrix

Once the convolved matrix is generated, then the rectified linear unit (ReLu), which is a non-linear activation function, is applied to it. This ReLu function replaces all the negative values with the value '0' and keeps all non-negative numbers the same. Pooling layer is the second layer under the hidden layer. These layers are used to reduce the dimension of the convolved matrix. The primary principle of these layers is to reduce the number of computations performed in image processing. In these models we are using max pooling layers. The maximum value for patches of a feature map is calculated and used to build a downsampled (pooled) feature map in the max pooling procedure. 2x2 matrices are generated at max pooling layer. The next process is flatting. Where the two dimension matrix(2D) is converted to a single array(1D) matrix. Next the output is fed to a fully connected layer. Fully connected layer is the last layer in the CNN model. Fully connected layers are used to compile all the data which was extracted in previous layers and form the final output.

All the libraries like Keras, TensorFlow, Mediapipe, OpenCV, time, Pygame, etc. have been imported and used to train the model capable of differentiating each gesture into a certain class. We have programmed each class to display its respective text and voice when it is predicted.

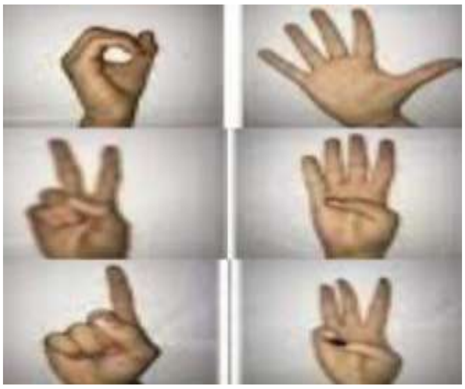
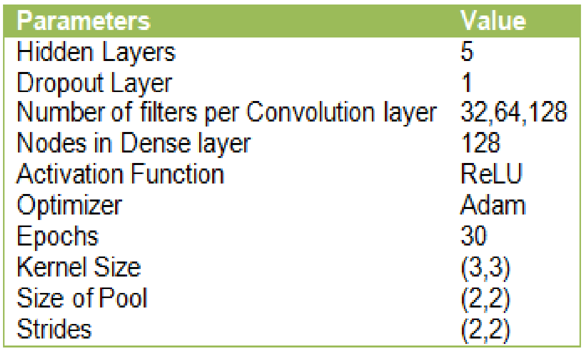


Fig 3. gestures used for training of CNN model

Table 1.Optimal hyperparameters for the CNN



Parameters	Value
Hidden Layers	5
Dropout Layer	1
Number of filters per Convolution layer	32,64,128
Nodes in Dense layer	128
Activation Function	ReLU
Optimizer	Adam
Epochs	30
Kernel Size	(3,3)
Size of Pool	(2,2)
Strides	(2,2)

Speech to gesture conversion:

The voice is given using the mobile application (voice boot). The mobile application is configured to connect with the Bluetooth device (HC05). There are two different modes available in the HC-05, such as data mode and command mode. The command is used to set the baud rate. The baud rate is set to 115200 by default, but we reduce it to 9600bps in order to transmit data and interface with the Raspberry Pi. frequency of the data transmission has been set, and the data will be transmitted smoothly. The voice data has now been saved in a non-readable format, which is stored in the variable created by the serial module. After interfacing HC05, we need to import a library called pyserial. These libraries are used to access the serial port in Raspberry. The voice from the microphone is transmitted to Raspberry Pi wirelessly by using the HC05 bluetooth module. The data which is presented at the serial port will be in the form of Unicode. This unicode will be converted to any data type, such as string or float. In my situation, I'm decoding with "utf-8," the Raspberry Pi's default encoding and the most widely used character encoding. After the conversion of the voice to the required format, i.e., string format, the text is displayed on the terminal. Whenever certain keywords appear in each text, we have programmed it to pop up its stored gesture image.[8]A few of the stored gesture images are shown below in Fig8.

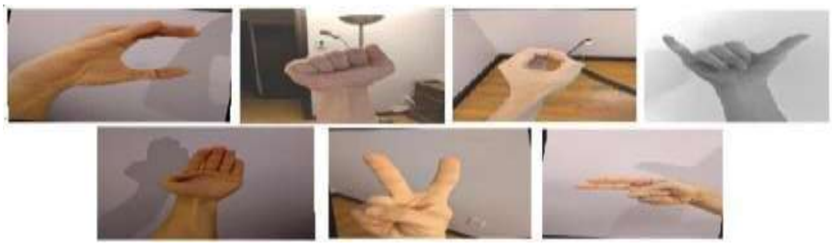
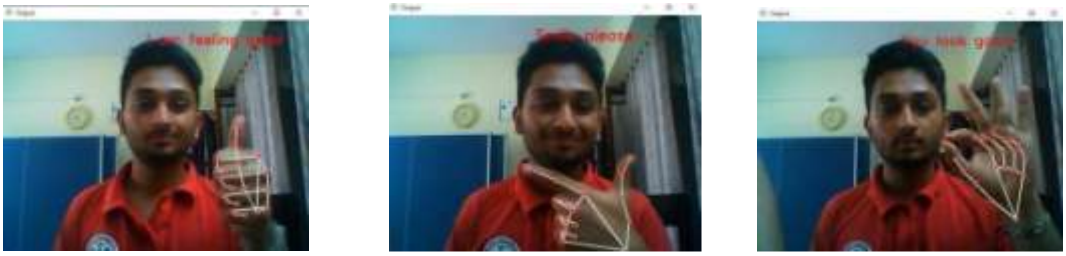


Fig 4. Few of the Gestures which appear when its respective text is called.

IV. Results

(a) Gesture to Speech conversion.

The findings of this study's experiments represent the model provided in this work can differentiate between multiple dominating features in photos at input side, as well as categorize diverse hand motions with. Fig 5(a) shows hand sign gesture images capturing and detecting text "I am feeling good," 5(b) "Smile please," and 5(c) "You look good".



(a) (b) (c)

Fig 5. Hand gesture detection

The model had an accuracy of 89.13 percent when it came to testing. Higher accuracy than the current one is difficult to achieve, but it is achievable by fine-tuning numerous hyperparameters and using adequate data preparation and augmentation approaches.

Figure 6 depicts the graph of the accuracy of the model on the y-axis with respect to the number of dataset applied to the CNN model.

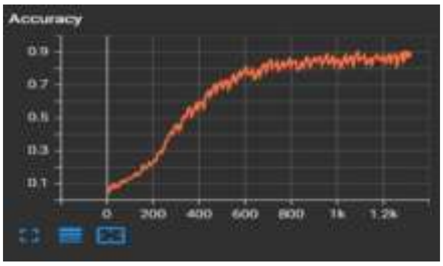


Fig 6. Plot of Accuracy vs Number of Dataset

Figures 7(a) and 7(b) show the output, which includes both text and gesture for the voice "Peace out" and "it is amazing," respectively.

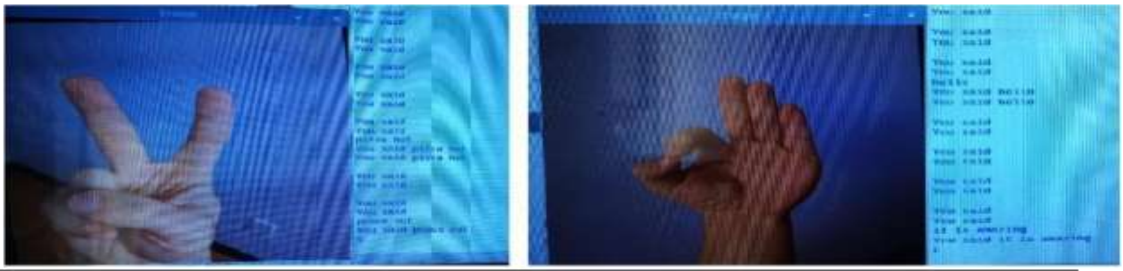


Fig 7..Voice detection and displaying of its text and gesture

V. Conclusion :

People who are deaf or dumb rely on language translators to communicate. They cannot, however, rely on interpreters on a daily basis, owing to the high expense and, as a result, the difficulty in obtaining and scheduling experienced interpreters. This technology, which is a two-way communication system for dumb and deaf people, will considerably improve the quality of life for impaired individuals. This is frequently a promising possibility; the necessary technology exists, and the prospective uses are interesting and desirable. When correctly calibrated and with suitable picture preprocessing, hand gesture recognition uses a convolutional neural network model to perform hand gesture identification on visual data. The photos and labels that were fed and trained within the model are utilized to compare the output when detecting pictures on live videos or static images. The system can provide two-way smart communication between the dumb and deaf community and the general public, effectively removing the communication barrier between them.

References

[1] Ishika Dhall, Shubham Vashisth, Garima Aggarwal, "Automated Hand Gesture Recognition using a Deep Convolutional Neural Network model" 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence).

[2] Prashant G. Ahire, Kshitija B. Tilekar, Tejaswini A. Jawake, Pramod B. Warale "Two Way Communicator Between Deaf and Dumb People And Normal People" 2015 International Conference on Computing Communication Control and Automation.

[3] Felix Zhan "Hand Gesture Recognition with Convolutional Neural Networks " 2019 IEEE 20th International.

[4] Houssem Lahiani, Mohamed Elleuch and Monji Kherallah, "Real Time Hand Gesture Recognition System for Android Devices", 2015, 15th International Conference on Intelligent Systems DeSign and Applications (ISDA). http://doi.org/10.1109/ISDA.2015.7489184.

[5] Rishabh Agrawal and Nikita Gupta, "Real Time Hand Gesture Recognition for Human Computer Interaction", 2016 IEEE 6th International Conference on Advanced Computing.

[6] Md. Mohiminul Islam, Sarah Siddiqua, and Jawata Afnan, "Real Time Hand Gesture Recognition using Different Algorithm Based on American Sign Language", ISBN.978-1- 5090-6004-7/17/ ©2017 IEEE.

[7] M. R. Abid, L. B. S. Melo, and E. M. Petriu, "Dynamic Sign Language and Voice Recognition for Smart Home Interactive Application", in Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA), May 2013, pp. 139–144.

[8] Na Zhao, Hongwu Yang "Realizing Speech to Gesture Conversion by Keyword Spotting" 2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP).