Sign Detection from Hearing Impaired

Manu E S L[1], Hariprasad E[2], Mohammed Fahd N[3], Premalatha K[4]

[123]*Department of Electrical and Engineering, Kumaraguru College of Technology[autonomous], Coimbatore, India*
[4]*Associate Professor, Department of Electrical and Engineering, Kumaraguru College of Technology[autonomous], Coimbatore, India*

---***---

Abstract - *Sign language is the mode of communication among the speech and hearing-impaired community. It is hard for common people who are not familiar with sign language to communicate without an interpreter. This paper focuses on transcribing the gestures from sign language to a spoken language which is easily understood by the listening. Alphabets and words from static images are among the motions that have been translated This becomes more important for the people who completely rely on gestural sign language for communication when they try to communicate with a person who does not understand the sign language. Most of the systems that are currently developed can recognize the gestures based on the skin color. This paper focuses on creating a filter that can detect the gestural features of the hand irrespective of the color. The detection is possible with the advent of computer vision and AI technologies like ML and DL. The systems currently employed use Vector machine Algorithms (ML) which has less accurate results. This paper focuses on using CNN (DL) for training of the AI model.*

Key Words: **Vector machine, Convolutional Neural Network (CNN), Deep learning, Computer Vision, Machine learning (ML)**

1.INTRODUCTION

People with hearing loss either by birth or any injury may not be able to talk. They converse with others via sign language. Usually sign language is not suitable to communicate with others. Only people in their close circle may be able to understand it. This paper focuses on providing an aid that helps hearing impaired people to communicate with the general public at ease.

[1] Using computer Vision, gestures can be detected at 93.17% accuracy. Gestures can also be Translated using capacitive touch sensor with 92% accuracy, also with some disadvantages. [2] Multiclass SVM could also be used to detect as fast as 0.017s as it is used for classification of multiple gestures trained. [3] As common people cannot always use an interpreter everywhere; we believe this will be an easier means of communication. The gestures are captured and converted into text messages. [4] The hardware could be a wrist band that takes video of the hands while gesturing. Analyzing the angles of fingers in each gesture, edge detection and threshold to the image should be done for appropriate results. [5] With the help of the camera, the signs are recognized and classified according to each class. The hand color differs from person to person, the images with dark background are involved. [6] Using CNN instead of SVM could predict the results with 6-7% increased accuracy. The recognition is done by CNN model and assisted by a microprocessor to point to the specific user message or action.

2. PROPOSED WORK

In this paper, we proposed a model that is built using Convolutional Neural Networks, that detects the sign given by the hearing-impaired people using OpenCv - a computer vision technology that helps in image processing and enables real-time recognition. We begin by preparing the dataset by taking pictures of the hand gestures and preprocessing them.

Having 26 classes of signs, 250 images per class were taken that sum up to a total of 6500 images in the dataset. Much more images were generated by data augmentation like rotation, flipping. Shear range, Cropping etc. The CNN model was built using Keras-a deep-learning API that works over Tensorflow specially built for AI practices. Fig-1 displays a high-level block diagram of the overall process. For the mathematical process in training, Numpy is used. The trained model is saved as a pickle file and tested against the test dataset, finally accuracy is obtained.

The saved model is deployed on a microprocessor that does the real-Time analysis of images. The AI detects a specific class of the sign and sends digital input to a microcontroller like an arduino and a text-to-speech converter like E-speak that gives speech output specific to the class detected on a speaker. The arduino then makes the processes like calling a trusted person in case of emergency or sending location via SMS using GSM and GPS. Also, the detected sign has a set of words or sentences that are shown on a display on the gadget.

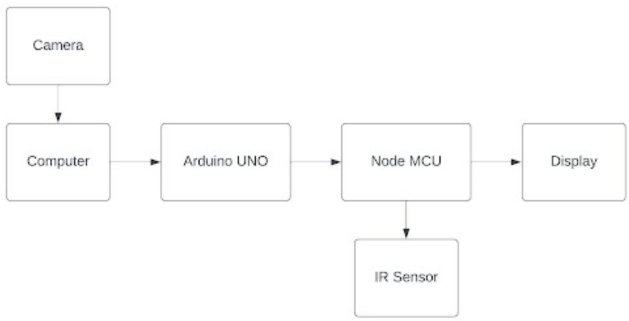


Fig -1: Block Diagram

3. IMPLEMENTATION AND RESULT ANALYSIS

3.1 COLLECTION OF IMAGE DATASET

The proposed system uses different images of faces with different angles and different poses. The images are taken by a specially developed code that captures images from the real time video. They are handled carefully and made into 26 classes. The taken images are then separated for test and training dataset. Training set contains 75% of the images taken and the 25% else goes to the test set. The images are categorized into 26 directories and labeled with alphabets for building purposes.

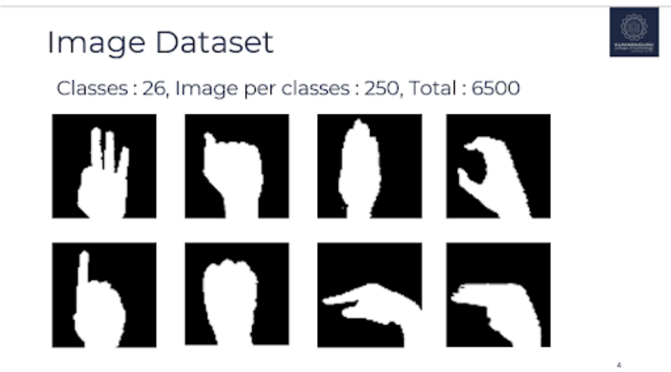


Fig -2: Dataset

3.2 PREPROCESSING OF IMAGES

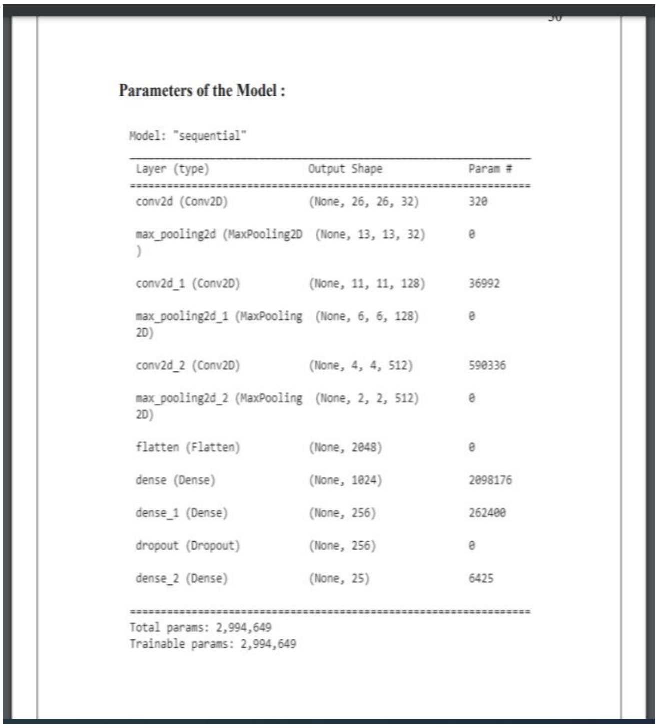
Pre-processing is the common operations made with images at the lowest level of abstraction. The aim of pre-processing is to improve the image data that suppresses unwanted distortions or to enhance some image features which are more important for further processing. The captured images in the RGB color model are converted into gray scale images by using HSV color transformation where; H – Hue (Dominant Wavelength), S – Saturation (Purity / shades of the color) and V – Value (Intensity). The HSV method is taken to get high efficiency output. Image segmentation is a process by which an image is partitioned into different regions. Image segmentation process includes background subtraction and foreground detection, followed by morphological operations of dilation and erosion.

3.3 BUILDING AND TRAINING OF THE AI MODEL

In the proposed architecture, the convolution was executed in the input data using a filter or kernel to create a feature map. However, convolution is done by sliding the input filter of 3x3. Finally, a map of various features and kept them together as the ultimate output of the convolutional layer. The output of the convolution passed through the activation functions such as Rectified Linear unit (ReLU) functions. Because it is nonlinear and provides the capacity to not stimulate all neurons at the same time, ReLU is the most extensively utilized activation function. After every level of convolution, a pooling layer was added within the CNN layer. The MaxPooling2D layer is a very basic layer in which no learning takes place and only dimension reduction occurs. It helps in reducing the number of learned parameters, thus reducing the computation and memory load.

The categorization was done in the fully-connected layer after the convolution and pooling layers. This fully-connected layer can only accept one-dimensional data. To convert 2D data to 1D data Python Flatten function is used. The flatten function is used to get a copy of a given array collapsed into one dimension. The fully connected layer is known as the Dense layer. The layer has 256 neurons. Each neuron in the flatten layer is connected to each neuron in the fully connected layer. 29 Dropout is used when a large feed forward neural network is trained on a small training set, it typically performs poorly on held-out test data which leads to “overfitting”. Fig 3 shows the no. of epochs trained, all the layers used dimensions of each layer, and total number of trainable parameters.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072



Parameters of the Model :

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 26, 26, 32)	320
max_pooling2d (MaxPooling2D)	(None, 13, 13, 32)	0
conv2d_1 (Conv2D)	(None, 11, 11, 128)	36992
max_pooling2d_1 (MaxPooling 2D)	(None, 6, 6, 128)	0
conv2d_2 (Conv2D)	(None, 4, 4, 512)	590336
max_pooling2d_2 (MaxPooling 2D)	(None, 2, 2, 512)	0
flatten (Flatten)	(None, 2048)	0
dense (Dense)	(None, 1024)	2098176
dense_1 (Dense)	(None, 256)	262400
dropout (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 25)	6425

Total params: 2,994,649
Trainable params: 2,994,649

Fig -3: Model Parameters

3.5 ANALYSIS OF THE PREDICTION

The trained model is now saved as a pickle file and then it is tested against a test set. Then the model is deployed in the local system and tested for real time recognition. Numpy and pandas helps greatly in analyzing the accuracy in graphical form. Also, Confusion matrix is obtained to visually to differentiate the class level precision. Fig-4 displays a confusion matrix where the prediction for each class is clear and nearly 1 with some classes having 0.95 prediction value.
FINAL ACCURACY: 99.25% accuracy on the test set

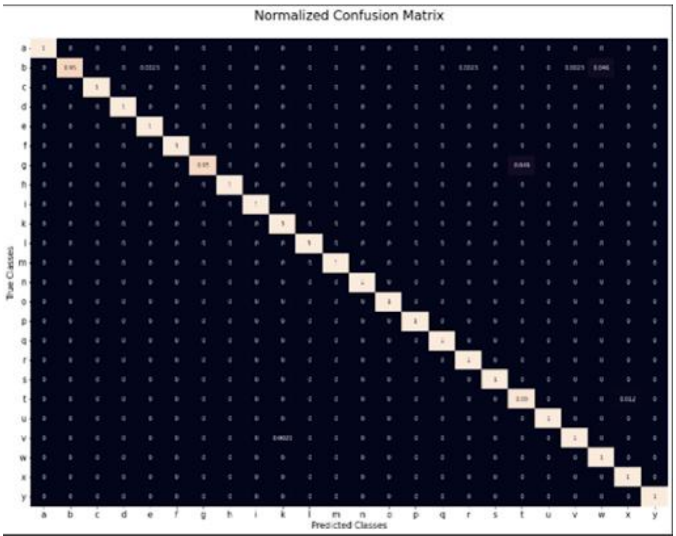


Fig -4: Confusion Matrix

4. TESTING AND RESULT EVALUATION
4.1 REAL-TIME TESTING

After the computational analysis of the prediction, the built model is tested in real-time. Fig-5 depicts a correct recognition of the class labeled ‘G’ then forwarded to a meaningful message - Good Morning that represents the class. The prediction accuracy depends on the resolution of the camera used and also the lighting of the surrounding area. The signs are to be shown in a green colored box, only there they are converted to grayscale and processed. For converting the colored to grayscale image, trackbars are used so that Hue, Saturation, and Value are adjusted and hands are distinguished from the background.

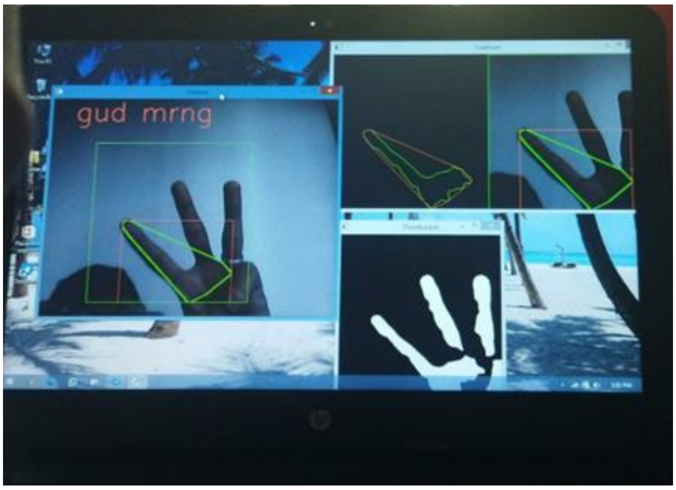


Fig -5: Sign Prediction

4.2. RESULTS

The model, the processing code, OpenCV code for video Processing are dumped into the hardware. Fig-7 shows that the accuracy was initially low and it attains a constant value near to 100 as the time increases. The predicted classes are then passed to a microprocessor and node MCU to make the expected outputs like emergency calling, location sharing and a speech output by speaker circuit.

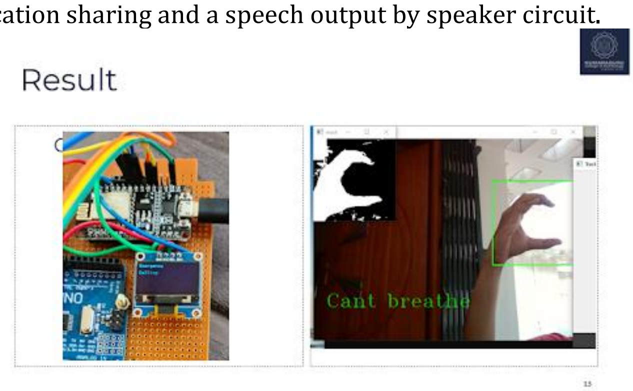


Fig -6: Accuracy Chart

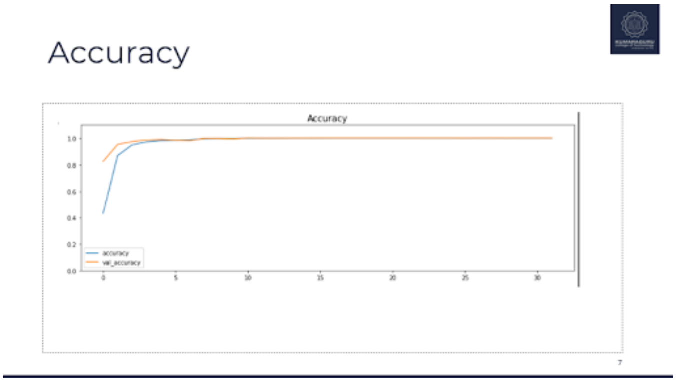


Fig -7: Accuracy Chart

5. CONCLUSION

This proposed system automatically detects the sign given by a speaking impaired person and converts it into a hearable voice. This project aims to develop a useful tool that uses gesture recognition for reducing the communication barrier between the deaf and dumb community and the normal people. Using this design, it is possible to convert hand gestures into speech which can be understood easily by normal people. The proposed system's concept provides more potential for future expansion. If more programming logic is introduced, more gestures could be incorporated. The proposed system successfully predicts the signs and some common words under different lighting conditions and different speeds. Accurate masking of the images is being done by giving a range of values that could detect the human hand dynamically. Also, at times of emergency, the exact location of the person wearing the gadget is sent to their trustable one and makes a phone call to a preferred person.

6. FUTURE SCOPE

This proposed idea could pave the way for the development of a wearable gadget that hearing impaired people can use to make sounds in emergency situations. Also, in future full conversation of the hearing-impaired people can be detected from Sign Language and converted to speech format. The same methodology can be used to make a tech that could automate homes using gestures. This can also be applied to other gadgets like changing the channels of television, control of other household products even in driving cars using Gestures.

7. REFERENCES

[1] Chen, F., Deng, J., Pang, Z., Baghaei Nejad, M., Yang, H., and Yang, G. (2018). Finger Angle-Based Hand Gesture Recognition for Smart Infrastructure Using Wearable Wrist-Worn Camera. Applied Sciences, 8(3), 369.doi:10.3390/app8030369.

[2] Dardas, N. H., and Georganas, N. D. (2011). Real –Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques. IEEE Transactions on Instrumentation and Measurement, 60(11), 3592–3607. doi:10.1109/tim.2011.2161140.

[3] Airò Farulla G, Russo LO, Pintor C, Pianu D, Micotti G, Salgarella AR, Indaco M (2014). Real-Time Single Camera Hand Gesture Recognition System for Remote Deaf-Blind Communication. Augmented and Virtual Reality, 35–52. doi:10.1007/978-3-319-13969-2_3.

[4] Abhishek K S, Qubeley L C F, Ho D. (2016). Glove-based hand gesture recognition sign language translator using capacitive touch sensor, 2016 IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC). doi:10.1109/edssc.2016.7785276

[5] V.Padmanabhan and M. Sornalatha. Hand gesture recognition and voice conversion system for dumb people International Journal of Scientific & Engineering Research, Volume Issue 5, May-2014.

[6] David J. Rios-Soria, Satu E. Schaeffer Sara, E. Garza-Villarreal, Hand-gesture recognition using computer-vision techniques Universidad Autónoma de Nuevo León (UANL) San Nicolás de los Garza, NL, México elisa.schaeffer@gmail.com.

[7] E K, Vyshakh C B and Shameer U Shahul. Gesture Controlled Speaking Assistance for Dump and Deaf National Conference On Discrete Mathematics & Computing (NCDMC-2017) e-ISSN: 2278-0661, p-ISSN: 2278-8727.