iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50962

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Real Time Sign Language to Text Conversion

Faiza Ansari[1], Nikhat Ansari[2], Maariyah Khanche[3], Iqra Shaikh[4]
MH Saboo Siddik College of Engineering , MD HAJI SABOO SIDDIK COLLEGE OF ENGINEERING, 8, MH Saboo Siddik Polytechnic Rd, Police Colony, Nagpada, Byculla, Mumbai, Maharashtra 400008, India

Abstract: ***This paper explores the potential to bridge the communication gap between those who speak sign language and those who do not understand it. Sign language is a visual language that is primarily used by people who are deaf or hard of hearing. It is a unique and complex form of communication that involves a combination of hand gestures, facial expressions, and body movements. However, for those who do not know sign language, it can be challenging to communicate with individuals who use it as their primary means of communication. By using advanced technology, we can convert sign language gestures into written or spoken language, allowing for effective communication between people who know sign language and those who do not. In this way, we can work to break down communication barriers and promote greater inclusion and accessibility for people who are deaf or hard of hearing.***
Keywords: Sign Language, Hearing impaired, Hand tracking.

I. INTRODUCTION

A sign language to text converter is needed to bridge the communication gap between people who are deaf or hard of hearing and those who do not know sign language.

Sign language is a unique form of communication that involves a combination of hand gestures, facial expressions, and body movements. It is an effective way for people who are deaf or hard of hearing to communicate with each other and with the hearing community. However, for those who do not know sign language, communication can be challenging, and it can lead to exclusion and isolation for people who are deaf or hard of hearing.

A sign language to text converter can help overcome these challenges by providing a means to translate sign language gestures into written or spoken language in real-time. This technology can promote inclusivity and accessibility for people who are deaf or hard of hearing and can help break down communication barriers. It is, therefore, essential to have sign language to text converters that are accurate, efficient, and widely available.

The development of a sign language to text converter using artificial intelligence (AI) and machine learning (ML) has the potential to revolutionize communication for people who are deaf or hard of hearing. By using complex algorithms and deep learning models, AI and ML can accurately recognize and interpret sign language gestures in real-time. This technology can also improve over time as it is trained on more data and can adapt to variations in signing styles and dialects. The result is a powerful tool that can convert sign language into written or spoken language, making it easier for people who do not know sign language to communicate with those who do.

The potential applications of such a system are vast and can have a significant impact on breaking down communication barriers and promoting greater inclusivity and accessibility for people who are deaf or hard of hearing.

American Sign Language (ASL) is a complete, natural language that has the same linguistic properties as spoken languages, with grammar that differs from English. ASL is expressed by movements of the hands and face. It is the primary language of many North Americans who are deaf and hard of hearing and is used by some hearing people as well. ASL is a language completely separate and distinct from English. It contains all the fundamental features of language, with its own rules for pronunciation, word formation, and word order. While ev

ery language has ways of signaling different functions, such as asking a question rather than making a statement, languages differ in how this is done [1].

American Sign Language is known as ASL. The deaf community in the United States and Canada uses it mostly as a visual language. ASL is not only a visual depiction of spoken English; it has its own distinct grammar, lexicon, and syntax. ASL is a sophisticated and intricate language that uses body language, facial emotions, and hand gestures to transmit meaning. It is an important tool for communication and gives deaf people a way to interact with others and express themselves.

ASL is taught in many colleges and institutions as a foreign language, and it is acknowledged as an official language in the United States.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

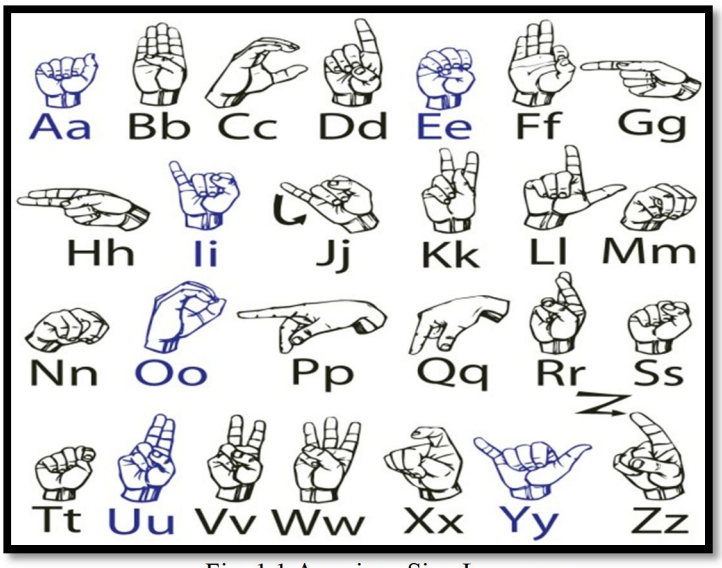


Fig. 1.1 American Sign Language

II. LITERATURE SURVEY

Jagdish L. Raheja et al. This paper describes a novel method of fingertips and centre of palms detection in dynamic hand gestures generated by either one or both hands without using any kind of sensor or marker. We call it Natural Computing as no sensor, marker or color is used on hands to segment skin in the images and hence user would be able to do operations with natural hand.

N.Tanibata et al. obtain hand features from a sequence of images. This is done by segmenting and tracking the face and hands using skin colour. The tracking of elbows is done by matching the template of an elbow shape. The hand features like area of hand, direction of hand motion, etc. are extracted and are then input to Hidden Markov Model (HMM).

D.Kelly et al. recognise hand postures used in various sign languages using a novel hand posture feature, eigen-space Size Function and Support Vector Machine (SVM) based gesture recognition framework. They used a combination of Hu moments and eigen-space Size Function to classify different hand postures.

H. K. Nishihara et al. (US patent, 2009), generate silhouette images and three-dimensional features of bare hand. Further, classify the input gesture by comparing it with predefined gestures.

Daniel Martinez Capilla used 8-dimensional descriptor for every frame captured by Microsoft Kinect XBOX 360 and compared the signs by dynamic time-warping (DTW).

III. METHODOLOGY

We put forward a method which uses deep convolutional networks for the classification of the images of the letters, digits and words in sign language.

We aim at representing features which will be learned by a technique known as convolutional neural networks (CNN) which contains four types of layers: convolution layers, pooling, subsampling layers, nonlinear layers and fully connected layers. This new representation is expected to capture various image features and complex non-linear feature interactions. Moreover, we have used a softmax layer to recognize signs.

CNN stands for Convolutional Neural Network. It is a type of deep learning algorithm used for image and video recognition, natural language processing, and other machine learning tasks. The basic idea behind CNNs is to use convolutional layers to extract features from input data, such as images. These layers apply filters to the input data, which allows the network to identify edges, shapes, and other patterns in the data.

The output from these convolutional layers is then passed to fully connected layers, which perform the actual classification or regression task. CNNs have proven to be very effective in computer vision tasks such as object detection, image segmentation, and facial recognition.

They are also commonly used in natural language processing applications, such as text classification and sentiment analysis. CNNs have been applied in a variety of industries, including healthcare, finance, and autonomous vehicles.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Convolutional neural networks are distinguished from other neural networks by their superior performance with image, speech, or audio signal inputs. They have three main types of layers, which are:

1) Convolutional layer
2) Pooling layer
3) Fully-connected (FC) layer

The convolutional layer is the first layer of a convolutional network. While convolutional layers can be followed by additional convolutional layers or pooling layers, the fully-connected layer is the final layer. With each layer, the CNN increases in its complexity, identifying greater portions of the image. Earlier layers focus on simple features, such as colors and edges. As the image data progresses through the layers of the CNN, it starts to recognize larger elements or shapes of the object until it finally identifies the intended object [2].

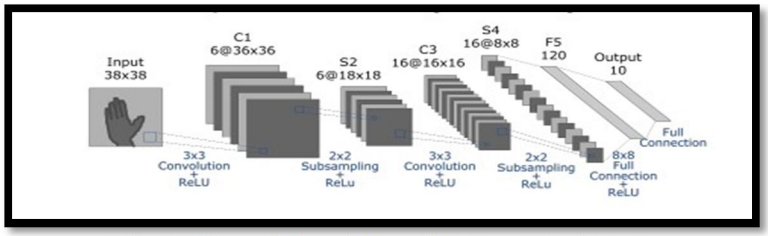


Fig 3.1 Convolutional Neural Network Architecture

Our approach was one of basic supervised learning by training the network with our own set of sign datasets. Our task is to use deep convolutional neural networks to classify letters and the digits 0-9 in ASL. The inputs were fixed size high-pixel images 50 by 50 dataset.

A. Dataset

At first, we trained and tested on a self-generated dataset of images we took ourselves. This dataset was a collection of 1200 images from multiple people for each alphabet and the digits 1-9. Since our dataset was not constructed in a controlled setting, it was especially prone to differences in light, skin color and other differences in the environment, so we also used a premade dataset to compare our datasets performance with.

B. Data pre-processing

For generating our own dataset, we captured the images for each sign, then removed the backgrounds from each of the images using background-subtraction techniques. When we initially split the dataset into two for training and validation, the validation accuracy showed to be high. However, the validation accuracy drastically decreased when we used two datasets from different sources i.e training on ours and testing on the premade and vice versa, since training on one dataset and validating on another was not yielding as accurate of results. We used the premade dataset for the different gestures to train the network which yielded the following results.

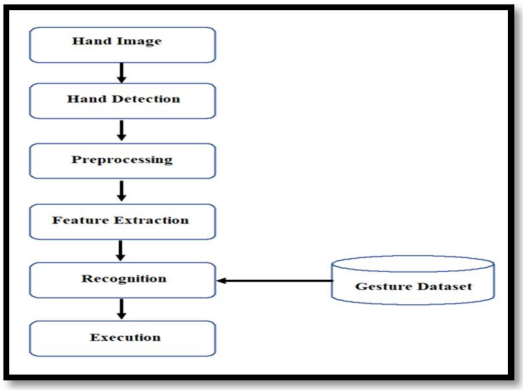


Fig 3.2 Process Flow

C. Skin Detection

For preprocessing the data, we first extract the skin region using the hsv color model.

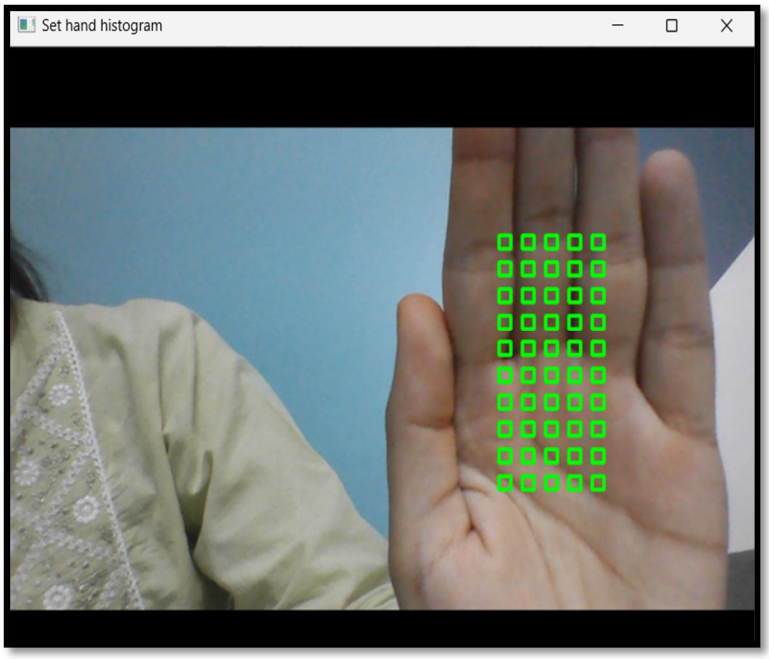


Fig 3.3 Skin Detection

D. Convolution and Filtering

We have applied a gaussian blurring median filter to generate a convolution of the input region of interest.

E. Contour Extraction and Thresholding

The outer most edges will be extracted as contours will be stored in the dataset folder for every gesture.

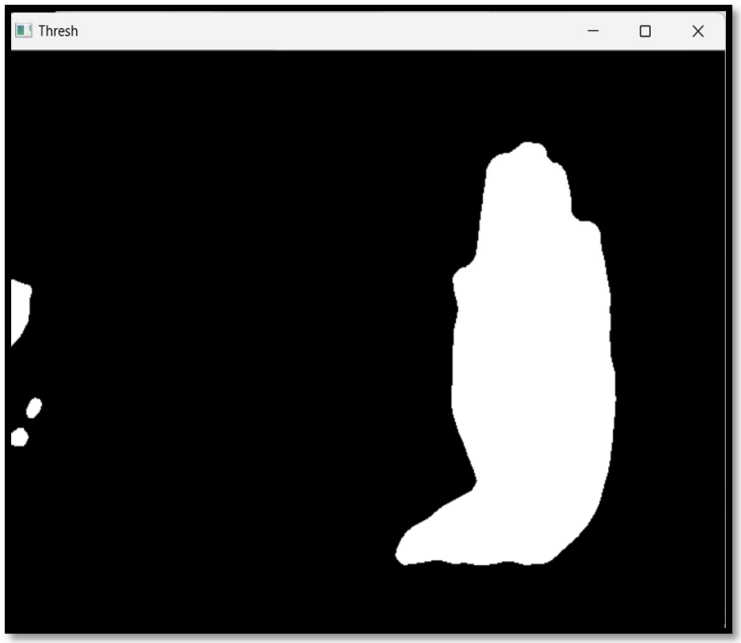


Fig 3.4 Contour Extraction and Thresholding

IV. SYSTEM DESIGN

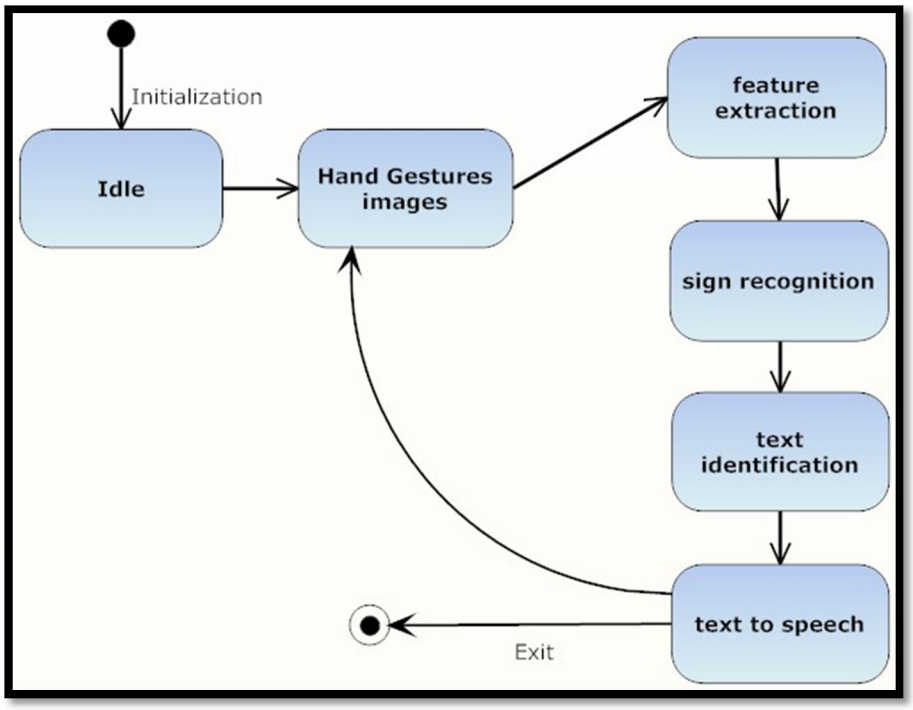


Fig 4.1 State Diagram

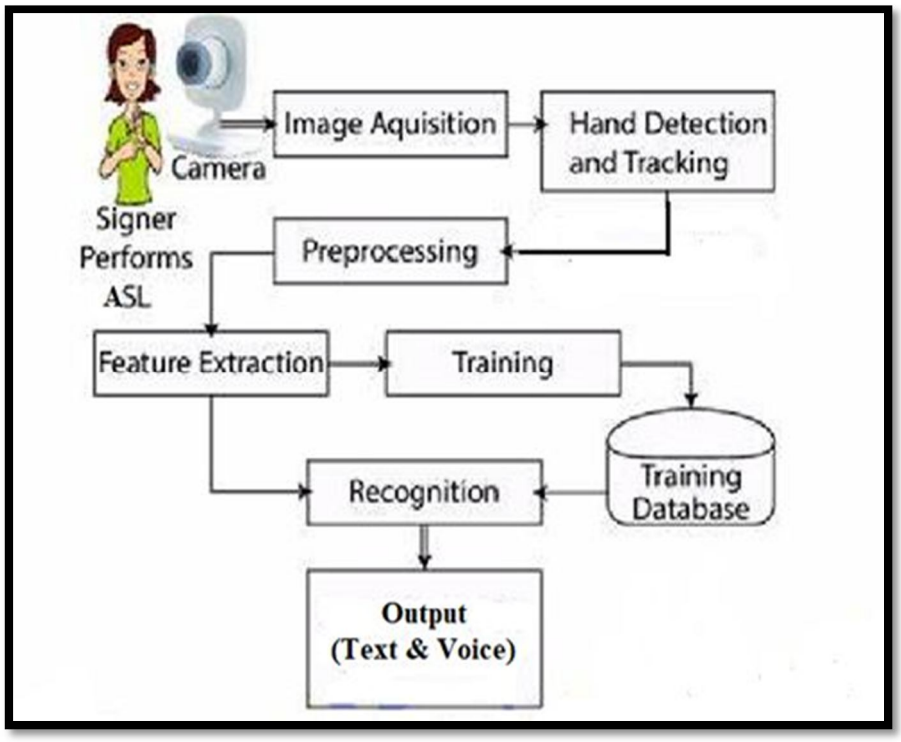


Fig 4.2 System Flow Diagram

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

V. RESULTS & DISCUSSIONS

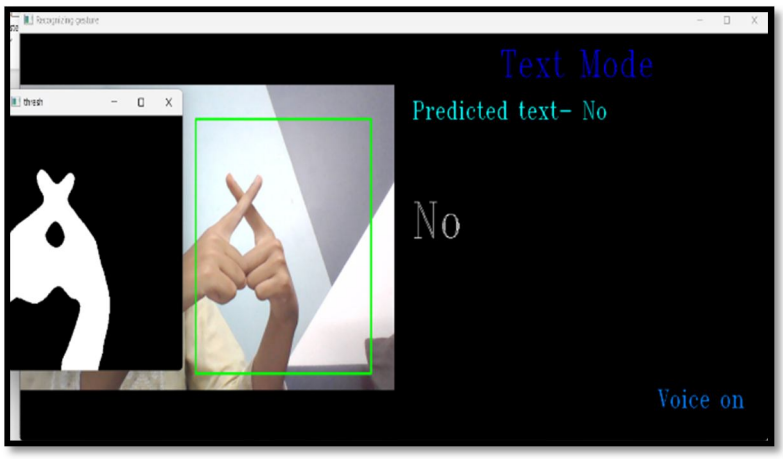


Fig 5.1 Sign Language “Help”

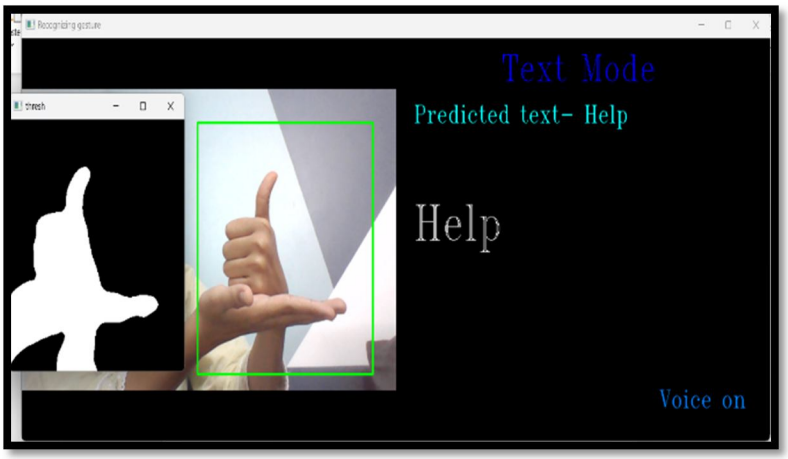


Fig 5.2 Sign Language “No”

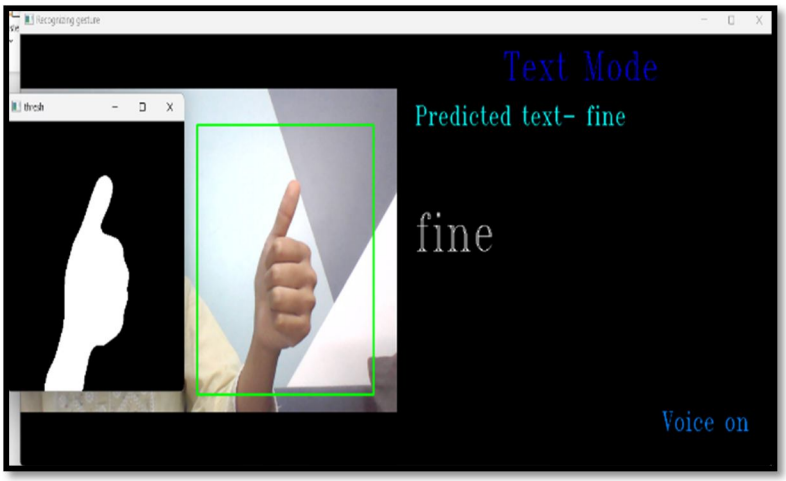


Fig 5.3 Sign Language “Fine”

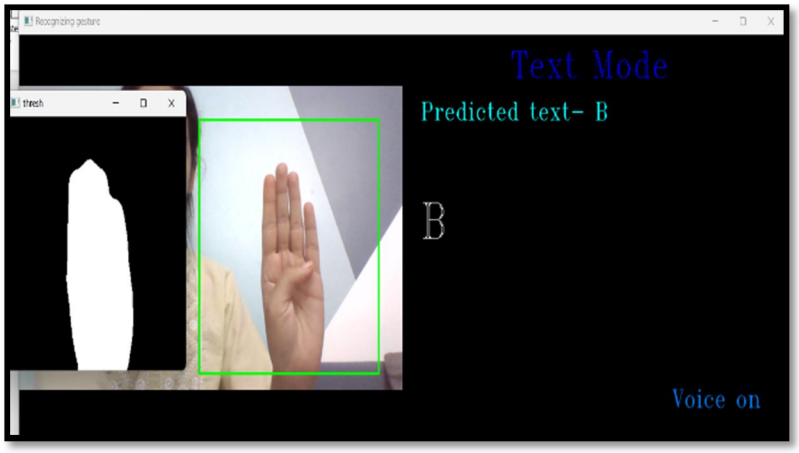


Fig 5.4 Sign Language “B”

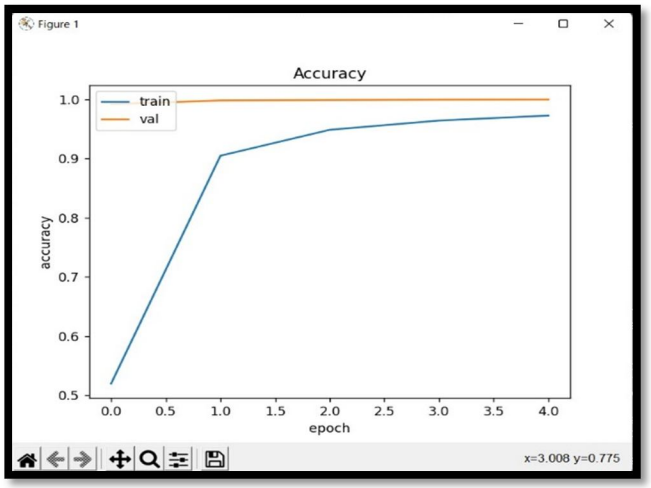


Fig 5.5 Accuracy Graph

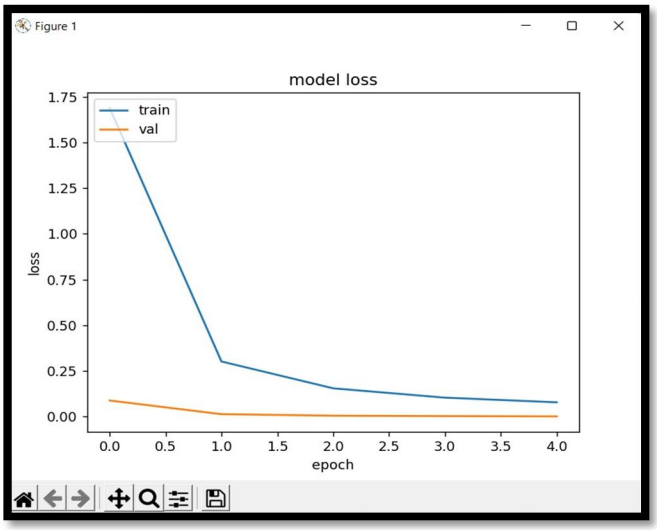


Fig 5.6 Model Loss Graph

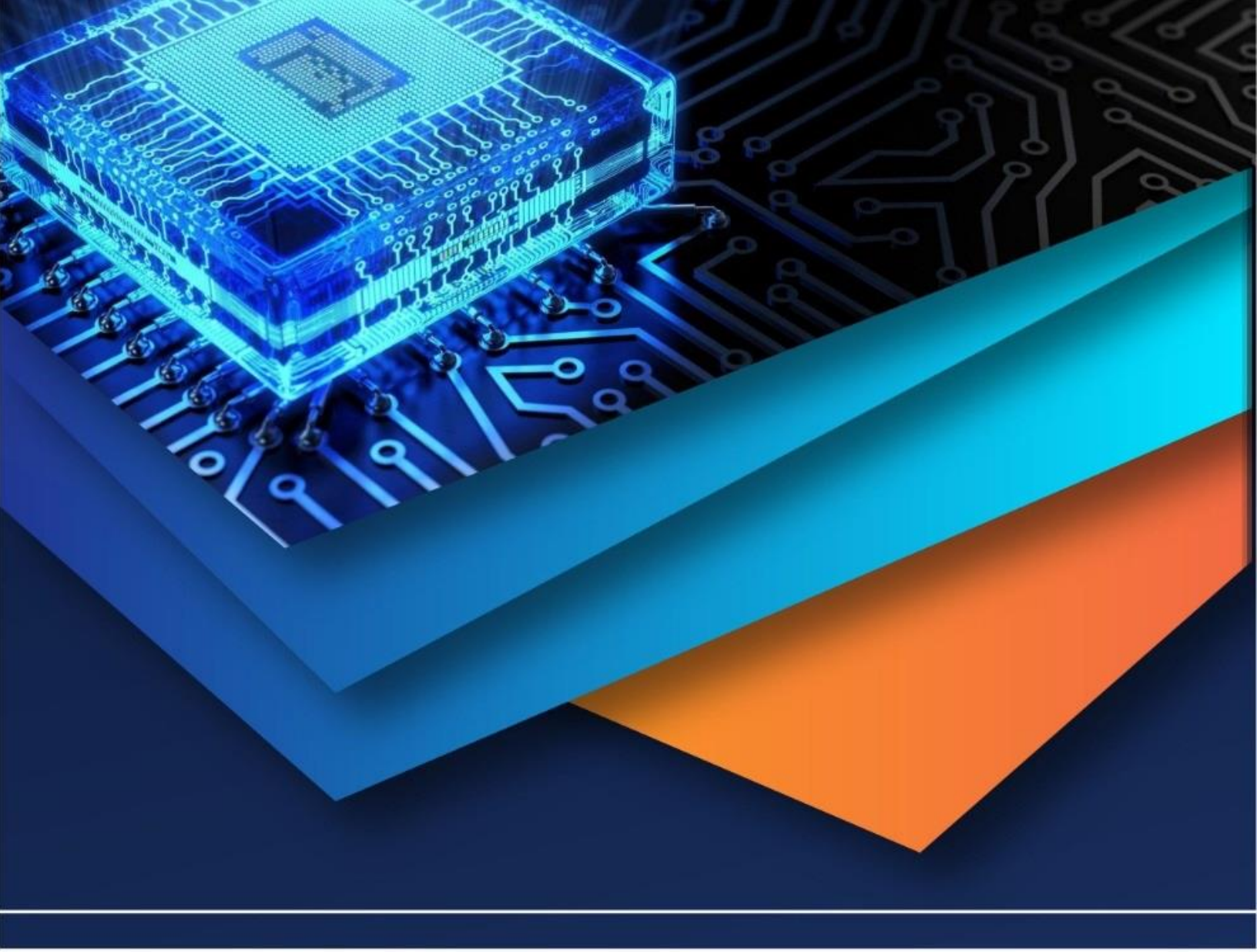
VI. CONCLUSION

This paper offers a user interface that supports Sign Language Recognition for simple communication with hearing-impaired users. The approach is not just applicable in a household setting but also in a public setting and is particularly useful for deaf and dumb persons in social situations. Based on the OpenCV toolbox, we will create a straightforward gesture recognizer and incorporate it into the Visionary framework.

REFERENCES

[1] https://www.nidcd.nih.gov/health/american-sign-language
[2] https://www.ibm.com/in-en/topics/convolutional-neural-networks
[3] J.L. Raheja, A. Chaudhary, K. Singal, "Tracking of Fingertips and Centre of Palm using KINECT", In proceedings of the 3 rd IEEE International Conference on Computational Intelligence, Modelling and Simulation, Malaysia, 20-22 Sep, 2011, pp. 248-252.
[4] N.Tanibata, N.Shimada and Y.Shirai, "Extraction of Hand Features for Recognition of Sign Language Words," International Conference on Vision Interface, pp.391-398, 2002.
[5] D.Kelly, J.McDonald and C.Markham, "A person independant system for recognition of hand postures used in sign language," Pattern Recognition Letters, Vol.31, pp.1359-1368, 2010.
[6] H. K. Nishihara et al., Hand-Gesture Recognition Method, US 2009/0103780 A1, date of filing Dec 17, 2008, date of publication Apr 23, 2009.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)