Global Scientific JOURNALS

GSJ: Vol 10,Issue 4,April 2022, Online: ISSN 2320-9186
www.globalscientificjournal.com

SIGN LANGUAGE RECOGNITION SYSTEM USING CNN AND NEURAL NETWORKS.

Hazel Lopes[1*], Anushka Agarwal[2], Janvi Bhalala[3], Khushit Trivedi[4]

[1*]*Dept. Computer Engineering, Universal College of Engineering, Vasai, India*
[2] *Dept. Computer Engineering, Universal College of Engineering, Vasai, India*
[3] *Dept. Computer Engineering, Universal College of Engineering, Vasai, India*
[4] *Dept. Computer Engineering, Universal College of Engineering, Vasai, India*

e-mail: hazel.lopes@universal.edu.in ,anushkaagarwal1234@gmail.com, janvibhalala02@gmail.com, khushittrivedi2000@gmail.com [9]

**Corresponding Author:hazel.lopes@universal.edu.in*

Available online at: http://www.ijcert.org

Received: 00/…./2021, Revised: 00/…./2021, Accepted: 18/May/2021, Published: 30/Aug/2021

Abstract:- To talk with a person with hearing or listening disability is always a major challenge. Hand gesture recognition system provides us an innovative, natural, user friendly way of interaction with the computer which is more familiar to human beings. Gesture recognition has a wide area of application including human machine interaction, sign language, immersive game technology etc. By keeping in mind the similarities of human hand shape, it aims to present a real time system for hand gesture recognition on the basis of detection of some meaningful shape based features. Sign language has indelibly become the ultimate panacea and is a very powerful tool for individuals with hearing and speech disability to communicate their feelings and opinions to the world. It makes the integration process between them and others smooth and less complex. However, the invention of sign language alone is not enough. The sign gestures often get mixed and confused for someone who has never learnt it or knows it in a different language. However, this communication gap which has existed for years can now be narrowed with the introduction of various techniques to automate the detection of sign gestures.In this study, the user must be able to capture images of the hand gesture using a webcam and the system shall predict and display the name of the captured image.The region of interest which, in this case is segmented hand gestures. It makes use of the Convolutional Neural Network(CNN) for training and to classify the images.

Keywords: Convolutional neural network, Deep learning, Gesture recognition,Sign language recognition, Hearing disability.

---

1. Introduction

American sign language is a predominant sign language Since the only disability D&M people have is communication related and they cannot use spoken languages hence the only way for them to communicate is through sign language. Communication is the process of exchange of thoughts and messages in various ways such as speech, signals, behavior and visuals. Deaf and dumb(D&M) people make use of their hands to express different gestures to express their ideas with other people. Gestures are the nonverbally exchanged messages and these gestures are understood with vision. This nonverbal communication of deaf and dumb people is called sign language.Using neural language and convolution systems, it was aimed to create a more efficient approach to the pre existing sign language systems.This report presents a report of a dual-cam first-person vision translation system for sign language using convolutional neural networks. The post is divided into three main parts: the system design, the dataset, and the deep learning model training and evaluation.

Sign language is a visual language and consists of 3 major components :

Fingerspelling	Word level sign vocabulary	Non manual features
Used to spell words letter by letter.	Used for the majority of communicatio n.	Facial expressions and tongue, mouth and body position.

In this project , the focus is on producing a model which can recognise Fingerspelling based hand gestures in order to form a complete word by combining each gesture. The gestures this project aims to train are as given in the figure 1.1

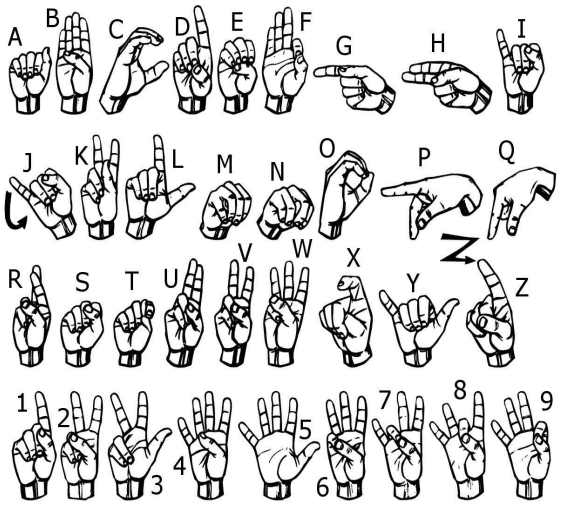

Fig 1:Datasets of ASL sign language

1.1 Convolution Neural Network

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 it will reduce the full image into a single vector of class scores.

Organization of the paper is organized as follows, Section I contains the introduction of the system in detail,Section II contain the literature survey, Section III contain proposed system, Section IV contain the results and discussion, section V contains the conclusion of the system, Section VI describes the future scope of the system, Section VII contain the references.

2. Related Work

A following research articles were selected for review:
In the research conducted by Mehreen Hurroo, Mohammad ElhamWalizad in their 2020 paper titled“Sign language recognition using convolutional neural networks and computer vision” proposed a system using low computing power to get maximum accuracy but the drawback here was hardware restriction leading to high cost. [1]

In another proposed system by Shailesh bachani, Shubham dixit, Rohin chadha, Prof. Avinash Bagul the concept of work was to convert easy words into gesture symbols and gestures into easy words, however this system was not compatible and accurate across the many sign language systems used globally. [2]

In other such recent paper by Siming He titled "Research of a Sign Language Translation System Based on Deep Learning" the paper proposes an automatic hand-sign language translator – a critical system for mute/deaf individuals the only drawback being high cost ineffectiveness and low accuracy.[3]

A system proposed by Dabre, Kanchan and Surekha Dholay. "Machine learning model for sign language interpretation using webcam images." A very basic model for recognition has been put forward which gave a rough idea how to get about the development stages of the project.[4]

Another interesting approach by Mohammed Elmahgiubi, Mohamed Ennajar, Nabil Drawil, and Mohamed Samir Elbuni in their paper "Sign language translator and gesture recognition" hinted on the design of hand locating algorithms based on deep learning, the feature extraction based on 3D CNN and the recognition algorithm.Even though this research is quite useful it is time complex making it redundant.[5]

In a paper proposed by Lionel Pigou(B), Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen titled " Sign language recognition using convolutional neural networks" it was initially hinted how convolutional neural networks can be used to accurately recognize different signs of a sign language.The limitation in this sye=stem was that it only catered to the American Sign language.(ASL)[6].

The reading and understanding of various literature research papers lead to a lot of insight as to the current technologies used and the problems faced by them. The analysis after referring to them include Sign Language Recognition Application Systems is developed in two steps, data acquisition and classification.

3. Methodology

Sign Language Recognition system is developed in two steps, data acquisition and classification. There are two data acquisition methods that are often used by researchers, Camera and Microsoft Kinect. The main advantage from using a camera is that it removes the needs of sensors in sensory gloves and reduces cost from building the system. As it is known that the camera is quite cheap and is available in almost all laptops. 6 are using high specification cameras because of the blur caused by web cameras. But even though it is a high
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 32 filter weights (3x3 pixels each).This will result in a 60 x 60 pixel image.
2nd Pooling Layer : The resulting images are downsampled again using a max pool of 2x2 and is reduced to a 30 x 30 resolution specification camera,it is still available in most smartphones. The disadvantage of using a web camera, or simply camera, is that good image pre-processing of obtaining the feature is needed.
The Microsoft Kinect is the other popular method used by researchers to acquire their data. Microsoft Kinect is getting more popular among researchers as it provides more data and it is needed by researchers. The advantages of using Kinect is that it provides the depth data of the video stream. The depth data is very useful as it can easily distinguish the background and the signer. Furthermore, it can be used to distinguish hands and body as the signer usually performs sign language by hands in front of their body. The disadvantage is that the Microsoft Kinect device is costly and it should be connected to the computer.

Implementation :

1. Whenever the count of a letter detected exceeds a specific value and no other letter is close to it by a threshold it prints the letter and adds it to the

current string(In this code values are kept as 50 and difference threshold as 20).
2. Otherwise it clears the current dictionary which has the count of images.
3.. 1st Densely Connected Layer : Now these images are used as an input to a fully connected layer with 128 neurons and the output from the second convolutional layer is reshaped to an array of 30x30x32 =28800 values. The input to this layer is an array of 28800 values. The output of these layers is fed to the 2nd Densely Connected Layer. By using a dropout layer of value 0.5 to avoid overfitting.
4. 2nd Densely Connected Layer : Now the output from the 1st Densely Connected Layer is used as an input to a fully connected layer with 96 neurons
5. Final layer: The output of the 2nd Densely Connected Layer serves as an input for the final layer which will have the number of neurons as the number of classes it classifies (alphabets + blank symbol).

Activation Function : It uses 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.

Pooling Layer : By applying max pooling to the input image with a pool size of (2, 2) 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. This layer "drops out" a random set of activations in that layer by setting them to zero. The network should be able to provide the right classification or output for a specific example even if some of the activations are dropped out.

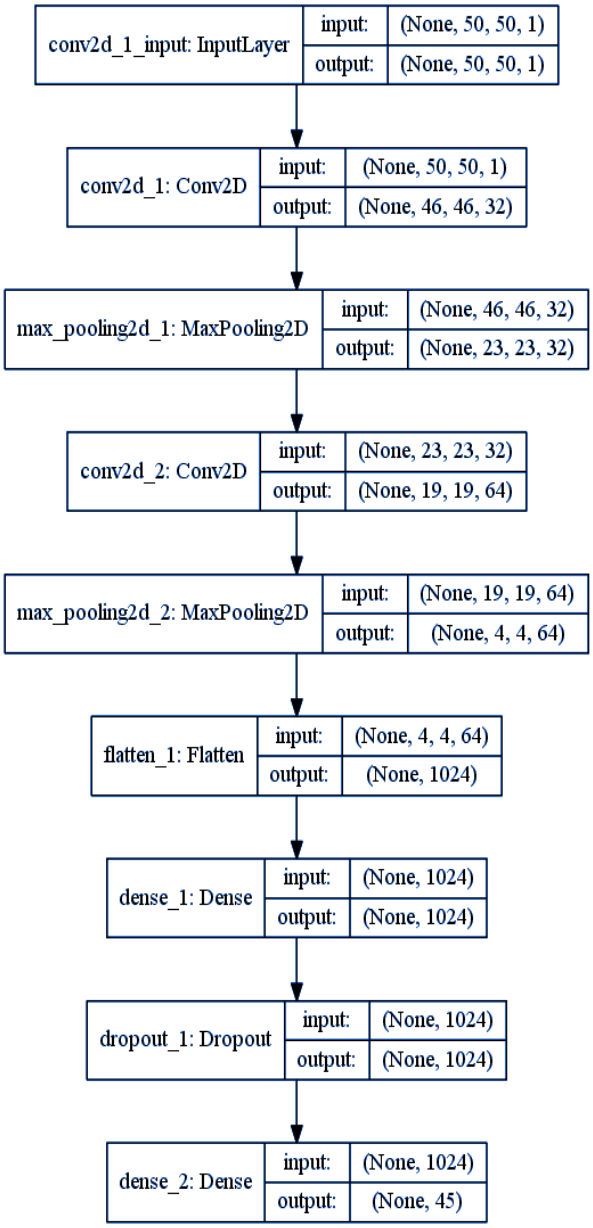

Fig 2: System Architecture of Sign language system

The system architecture for the system is shown in the figure 2 above.The Flow chart in Figure 3 demonstrates the approach taken to measure the system performance metrics.Testing each letter individually, where 50 iterations are applied on each letter. The frequency of recognizing any letter is stored in the table shown below. However, the recognition might be erroneous in the sense that the letter gesture is not recognized.

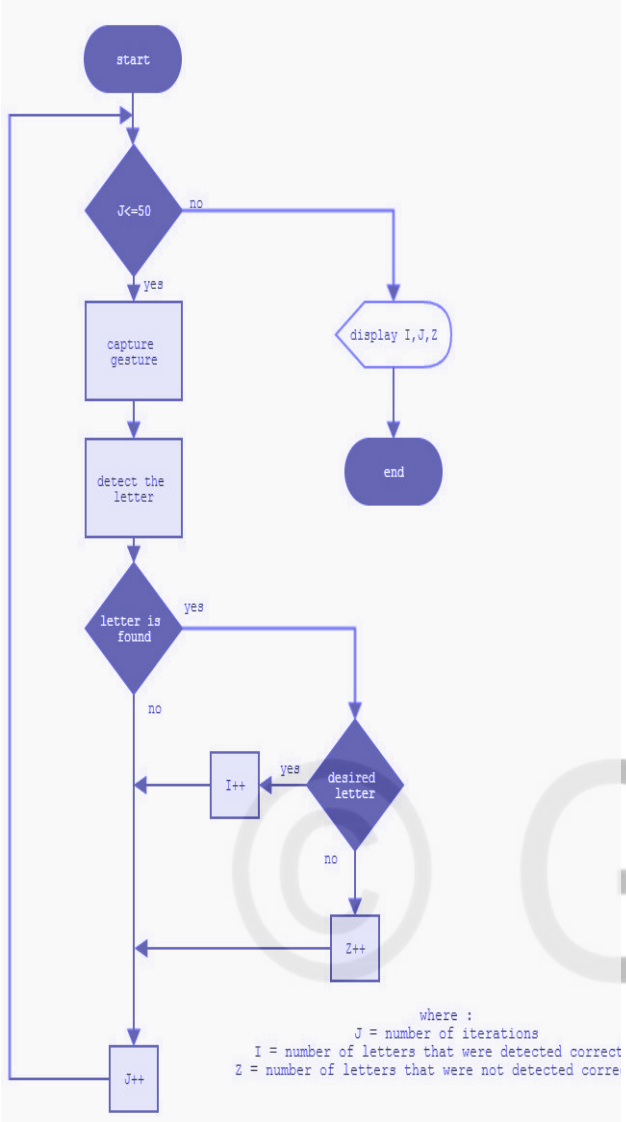

Fig.3:The testing procedure flowchart

raw images that matched the requirements. All the datasets that were available were the datasets in the form of RGB values. Hence,it was decided to create a data set. Steps this project followed to create the data set are as follows

Using Open computer vision(OpenCV) library in order to produce dataset. Then capture each frame shown by the webcam of the machine. In each frame, define a region of interest (ROI) which is denoted by a green bounded square as shown in the image.

Finally by applying a gaussian blur filter to the image which helps us extract various features of the image. The image after applying gaussian blur looks like the figure shown below.

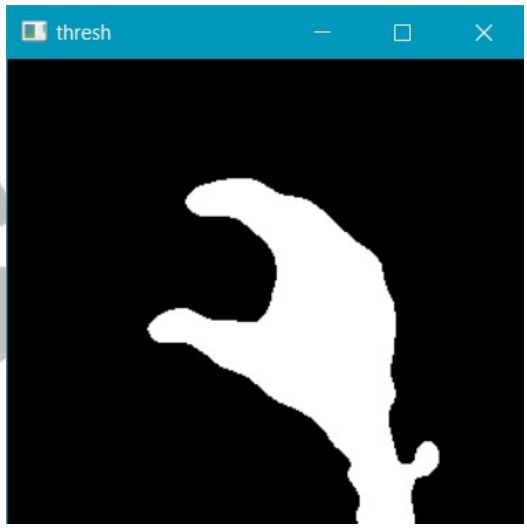

Fig 4:Image After Applying Gaussian Blur

4. Results and Discussion

This section depicts the "Sign Language Recognization System using CNN and Neural Networks". The screenshots below are from the system we built to make it easier for the people with disabilities to be more independent. This system is convenient to use and quite intuitive because it is both cost-effective and efficient.
The system is a vision based approach. All the signs are represented with bare hands and so it eliminates the problem of using any artificial devices for interaction.

Data Set Generation
For this project, finding pre-made datasets was difficult as we couldn't find datasets in the form of

After assembling the hardware and loading the desired code into our chip, iterative tests were done to make sure that the desired requirements have been met.
The accuracy and error rates are calculated using the following equations:

$$\textbf{Accuracy\%}=\frac{Detected\ right}{No\ of\ iterations}*100$$

The recall, f1 score and support can be calculated by :

$$\textbf{Precision} = \frac{True\ positive}{Actual\ Results}$$

$$\textbf{Recall} = \frac{True\ Positive}{Predicted\ Results}$$

F1 score= $\frac{2*(precision*recall)}{precision+recall}$

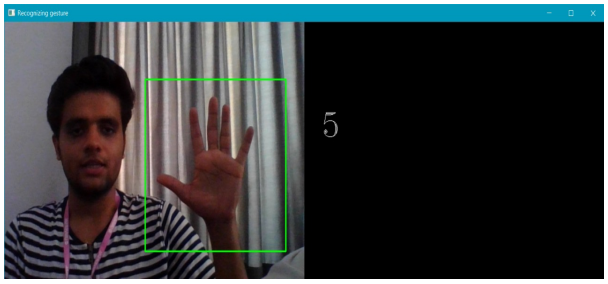

Fig 5:Gesture Recognition By CNN Model

The system works on two modes: Text mode and Calculator mode which can be switched from one mode to another by entering "t" or "c" on the keyboard..There are 10 operators to help in the implementation.

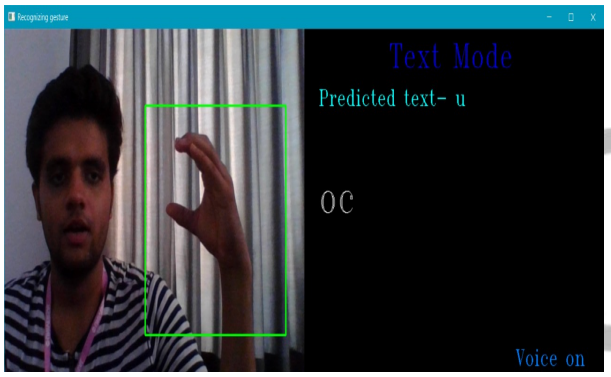

Fig 6:Gesture Recognition in Text Mode

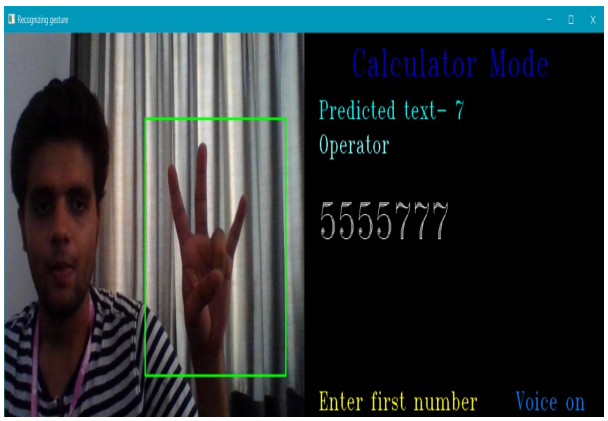

Fig 7: Gesture Recognition in Calculator Mode

Table I depicts two types of recognition: not detected and detected wrong. The second error type exposes the ambiguity issue in the recognition process. This ambiguity happens when
two or more gestures have similar finger positions or moves– as a result flex sensors can not do much in such cases. For example, some letters are mistaken for others like the letter "v"and "p" where the values of their flexes are relatively similar.

Table I:Accuracy and Error Values.

Title	Precision	Recall	F1 score	Support
0	0.99	0.99	0.99	195
1	1.00	1.00	1.00	205
2	0.99	0.99	0.99	198
3	0.99	0.99	0.99	231
4	0.99	0.99	0.99	193
5	0.99	0.99	0.99	193
6	0.99	1.00	0.99	214
7	1.00	1.00	1.00	209
8	0.99	0.99	0.99	185
9	1.00	1.00	1.00	212
10	0.99	1.00	0.99	199
11	0.99	0.99	0.99	183
12	0.99	0.97	0.98	188
13	0.97	0.99	0.98	195
14	1.00	1.00	1.00	203
15	0.99	1.00	1.00	226
16	1.00	1.00	1.00	187
17	0.98	1.00	0.99	191
18	1.00	0.99	0.99	191
19	0.98	0.99	0.99	172
20	0.99	0.99	0.99	190
21	1.00	0.98	0.99	222
22	1.00	1.00	1.00	210
23	0.99	1.00	1.00	198

24	1.00	1.00	1.00	210
25	1.00	0.95	0.98	211
26	1.00	0.99	0.99	204
27	0.99	0.99	0.99	200
28	0.99	1.00	0.99	198
29	1.00	0.99	0.99	188
30	0.99	0.99	0.99	197
31	0.99	1.00	0.99	184
32	0.99	1.00	0.99	198
33	1.00	0.99	1.00	226
34	0.99	0.99	0.99	200
35	1.00	1.00	1.00	201
36	1.00	1.00	1.00	189
37	0.98	1.00	0.99	201
38	1.00	1.00	1.00	189
39	0.99	0.98	0.99	201
40	1.00	1.00	1.00	217
41	1.00	1.00	1.00	211
Accuracy			**0.99**	**8400**
Macro Avg	**0.99**	**0.99**	**0.99**	**8400**
Weighted Avg	**0.99**	**0.99**	**0.99**	**8400**

Since this approach is not hardware dependent it is very cost efficient and the easy to use interface makes it efficient for all people to use.

5. Conclusion and Future Scope

Expected requirements and level of performance of such systems are addressed in this section..The software part of this system is extensively elaborated to include simple system initialization and recognition algorithms. The report addresses the challenges of identifying ambiguous measurements and proposes respective technical solutions.

With the continuous breakthrough of neural networks in artificial intelligence, computer vision and other related fields, neural networks have brought dynamic new methods to the study of sign language recognition based on vision. Starting from the task of common sign language word recognition, and focusing on the topic of sign language locating and sign language recognition based on neural network, this paper discusses the design of hand locating algorithm based on deep learning, the feature extraction based on 3D CNN, and achieves better recognition results than other methods on common vocabulary datasets.

Nowadays, applications need several kinds of images as sources of information for elucidation and analysis. Several features are to be extracted so as to perform various applications. When an image is transformed from one form to another such as digitizing, scanning, and communicating, storing, etc. degradation occurs. Therefore,the output image has to undertake a process called image enhancement, which contains a group of methods that seek to develop the visual presence of an image.Image enhancement is fundamentally enlightening the interpretability or awareness of information in images for human listeners and providing better input for other automatic image processing systems. Image then undergoes feature extraction using various methods to make the image more readable by the computer.

The current solution is the best because it correctly identifies the gestures by extracting the features and correctly recognizing the sequences in which the gesture takes place.It is evident from the experimental results that the system has the potential to help targeted individuals and communities. Especially that the system was able to recognize most of the letters (20 out of 26), and get to an average accuracy of 96%.

Future scope of this project would be to add the remaining letters to the system and better the system performance.Sign language recognition system is a powerful tool to test an expert's knowledge in edge detection.The intent of

convolution neural network is to get the appropriate classification of all gestures.

Also this system can be collaborated with the various high technological HMI softwares like Siri,Bixby and Google Assistant to make it more mobile to the general public.

Acknowledgement

We would like to thank our Campus Director Dr.J.B.Patil, our HOD Dr. Jitendra Saturwar and our project guide Mrs.Hazel Lopes for guiding us throughout the project and to bring the best out of us.

6. References

1. Mehreen Hurroo , Mohammad Elham, 2020, Sign Language Recognition System using Convolutional Neural Network and Computer Vision, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 12 (December 2020),

2.Shailesh bachani, Shubham dixit, Rohin chadha, Prof. Avinash Bagul,2020,SIGN LANGUAGE TO TEXT AND SPEECH CONVERSION USING CNN International Research Journal of Modernization in Engineering Technology and Science(IRJMETS) Volume:03/Issue:05/May-2021

3.S. He, "Research of a Sign Language Translation System Based on Deep Learning," 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), 2019, pp. 392-396, doi: 10.1109/AIAM48774.2019.00083.

4.Dabre, Kanchan and Surekha Dholay. “Machine learning model for sign language interpretation using webcam images.” 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA) (2014): 317-321.

5.M. Elmahgiubi, M. Ennajar, N. Drawil and M. S. Elb Uni, "Sign language translator and gesture recognition," 2015 Global Summit on Computer & Information Technology (GSCIT), 2015, pp. 1-6, doi: 10.1109/GSCIT.2015.7353332.

6.Pigou, Lionel, et al. “Sign Language Recognition Using Convolutional Neural Networks.” Lecture Notes in Computer Science, Springer, 2015, pp. 572–78, doi:10.1007/978-3-319-16178-5_40.

Authors Profile

Mrs. Hazel Lopes Currently working as an Assistant Professor of the computer science department at Universal College of Engineering, Mumbai.Her area of interests are Machine Learning and Artificial Intelligence.

Ms. Anushka Agarwal is currently pursuing B.E in Computer Engineering from Universal College of Engineering affiliated to University of Mumbai. Her area of interest is Automation and Data Visualization.

Ms. Janvi Bhalala is currently pursuing B.E in Computer Engineering from Universal College of Engineering affiliated to University of Mumbai. Her area of interest is Artificial Intelligence.

Mr. Khushit Trivedi is currently pursuing B.E in Computer Engineering from Universal College of Engineering affiliated to University of Mumbai. His area of interest is Augmented Reality and neural networks.