IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 4 (April. 2019), ||S (II) || PP 17-25

Sign Language Interpretation using Deep Learning

Jessica Dias[1], Dishita Patil[2], PraptiRaut[3], Malvina Lopes[4]
[1](Department of Information Technology, St. John College of Engineering and Management, India)
[2](Department of Information Technology, St. John College of Engineering and Management, India)
[3](Department of Information Technology, St. John College of Engineering and Management, India)
[4](Department of Information Technology, St. John College of Engineering and Management, India)

Abstract: Sign language is a language that Deaf people use to communicate with other normal people in the community. Although the sign language is known to hearing-impaired people due to its widespread use among them, it is not known much by other normal people. In this project, we have developed a sign language recognition system for people who do not know sign language, to communicate easily with hearing-impaired people. This project is built to interpret American Sign Language and also provides a complete overview of deep learning-based methodologies for sign language recognition. This will benefit deaf and hearing-impaired people by offering them a flexible interpreting alternative when face-to-face interpreting is not available. And the main purpose of our project is to develop an intelligent system which can act as a translator between normal people and deaf or dumb people, and can be the communication path between people with speaking deficiency and normal people with both effective and efficient way.

Keywords: American Sign Language, Deep learning, Real-Time, Convolution Neural Network

---

Date of Submission: 27-03-2019 Date of acceptance: 12-04-2019

I. INTRODUCTION

Very few people understand Sign language. Deaf people are usually deprived of normal communication with other normal people in the society. It's been observed that the deaf people find it really difficult at times to interact with normal people with their gestures, a very few of those are recognized by most people. Since people with speaking deficiency can't talk like normal people so they have to depend on some sort of visual communication in most of the time. Sign language is a language that provides visual communication and allows individuals with hearing impairments to communicate with other normal individuals in the community. Hence, the need to develop automated systems capable of translating sign languages into words and sentences is becoming a necessity [8]. And the availability of such translator is really limited, expensive and does not work throughout the life period of a deaf person. So, the solution is that computerized system is most relevant and suitable for translating signs expressed by deaf people into text and voice. The Image Processing method is used for better extraction of features from input images, that should be invariant to background data, translation, scale, shape, rotation, angle, coordinates, movements, etc. Also, Neural Network Model is used to recognize a hand gesture in an image. [3]. Deep Learning is a relatively recent approach to machine learning that involves neural networks with more than one hidden layer. Networks based on deep learning paradigms enjoy more biologically inspired architecture and learning algorithms, in contrast to conventional feed-forward networks. Generally, deep networks are trained in a layer-wise fashion and rely on more distributed and hierarchical learning of features as it is found in the human visual cortex; The data used in this work are obtained from a surrounding environment and contain different hand gestures for recognition. In order not to over-bias learning, the image samples of hand gestures are used for training and testing the designed networks.

The dataset used, contains a sampled image set of all-American Sign Language (ASL) alphabets Aa-Bb signs and 1-10-digit signs can be observed in Figure (1).The data in its raw form is provided as a pixel to pixel intensity [0-255] class-wise distributed XLS files and data preprocessing steps included conversion of the mentioned data to image format using PNG format 64*64 grayscale images. The collected data is separated into training and testing data, 80% data is given for training and 20% data is given for testing. After splitting data, the model is created as sequential network and started fitting process. Fitting process ran through all train data. After, training step, the model and weights and neural network loaded into real-time recognition algorithm. The algorithm consists of two parts that run simultaneously for better accuracy 1) is extracting hands bound feature points, 2) is classifying hand image with convolutional neural network. When there are similar hand signs, the decision will be made according to those steps results.

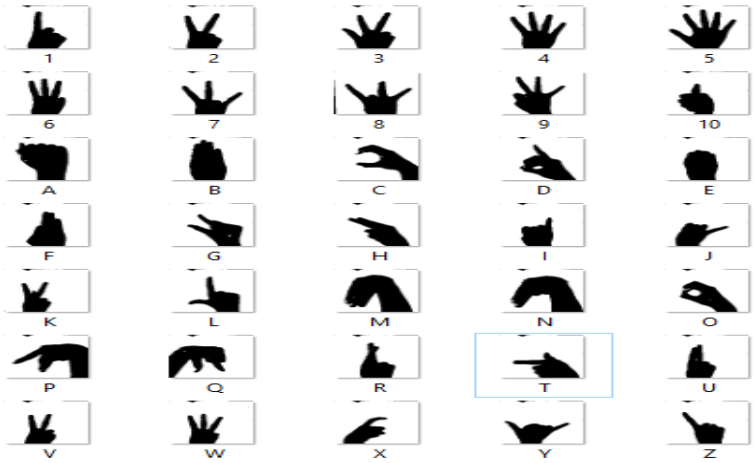

Figure 1. American Sign Language Alphabets and Digits

Machine Learning has a very prevalent subcategory called as deep learning,because of deep learning gives high level of performance over the data. Deep learning use to categorize images to build a convolutional neural network (CNN). The Keras library in Python isuse to build a CNN.Pixels in images are generally interrelated. Images are seen by computers by pixel value. For example, pixels in images may indicatesome pattern. This pixels value use by convolution to recognize images.And matrix of pixels multiplies with a filter matrix and calculates up the multiplication values by convolution. And then convolution moves to next pixel and follow same process as above until it completes all the pixels presents in an image

II. PROPOSED METHODOLOGY

The research done in Sign Language Recognition field are mostly done using glove-based system. in the glove-based system, sensors such as potentiometer, accelerators, etc. are attached to each of the finger. Based on their readings the corresponding alphabets is displayed. And over the years, advanced glove devices have been designed such as the Sayre Glove, Dexterous Hand Master and Power Glove.

The main problem faced by this gloved based system is that it has to be recalibrated every time a new user puts the glove so that the fingertips are identified by the Image Processing unit. Since most of the gloves are made solely in very few totally different sizes, the simple factor is to use custom created gloves that dead fits the user's hand. And makes a selected glove dead fitting just only for one specific person. Also because of frequent use the glove will be broken. The connecting wires restrict the freedom of movement. It also complexes to implement and hardware requirement is also more. This system is not cost effective.

III. PROPOSED ARCHITECTURE

The system is designed to capture an input sign image which will undergoes various image processing techniques. Firstly, an input image will be converted from RGB to Grayscale. Noise present in the input image will be removed for better accuracy of the input gesture. Further the hand part is detected from the image and hand gesture will be removed from the obtained image. Then this processed image is then compared with the trained model.

Phase-1: In this phase, a User Interface is developed in which a user can capture the image from webcam. These captured images are stored in the image input folder. Also, the hand gesture images are collected from the surrounding environment which can be used to feed CNN model for training. This image collection includes alphabets hand gestures and 1-10-digit hand gesture images. There are 1000 images per class i.e. for each alphabet and digits (1-10).

Phase-2: In this phase, the images collected are given for training the CNN model. In this the images are converted into grayscale and the images are feed to CNN model for training. In which 80% of data given for training and 20% for data given for testing. In training output the model produce model_path.h5 file which stores summary of the training step. This file will be used by model for prediction of alphabets and digits.

Phase-3: In this phase, the input image from the user is given to the CNN model for prediction, in which the input images and the images stored in the CNN model are compared. Based on the comparison the CNN model gives an output in text or audio format

Sign Images:

Hand gesture images are collected from surrounding environment. These are the different hand gesture sign images collected for training the CNN model for the recognition for signs.

Gray Scaling:

Gray-scaling is the process of converting a continuous-tone image to an image that a computer can be able to manipulate. While gray scaling is an improvement over monochrome. Here the collected data undergoes color conversion where the image is converted into grayscale. The data collected earlier spllitted into training and testing data.

Training Data:

For evaluating models, separating data into training and testing sets is an important part. Typically, when you separate a data set into a training set and testing set, most of the data is used for training, and smaller portion of the data is used for testing. This data is used to fit and tune your models. For images of training data collection consists of 1000 images of each hand gesture. This phase trained the model by extracting the features present in given training dataset.

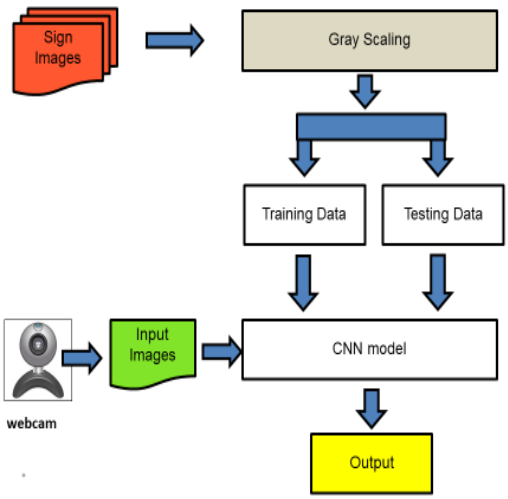

Figure 2. Architecture Diagram

Testing Data:

A subset to test the trained model. A testing data is a dataset that is independent of the training dataset and follows the same probability distribution as the training dataset. Also, a set of observations used to evaluate the performance of some model using performance metric. And is important that no observations from the training set are included in the test set. These are put aside as "unseen" data to evaluate your models. For images of testing data collection consists of 250 images of each hand gesture.

Neural Network:

Neural Network is a human brain algorithm that is designed to recognize pattern in numerical datasets. Image, text audio, video, etc., are the examples of real-world data; that needs to transform into numerical vectors to use neural networks. This network is composed of different layers and a layer is made of multiple nodes. The mapping of the input to the output is performed by some activation function. The goal of neural network is to approximate some function 'f'. A task of simple classifier function y=f(x), which maps the input data x to a class y, while the neural network identifies β, that results in proper function, y=f (x, β).

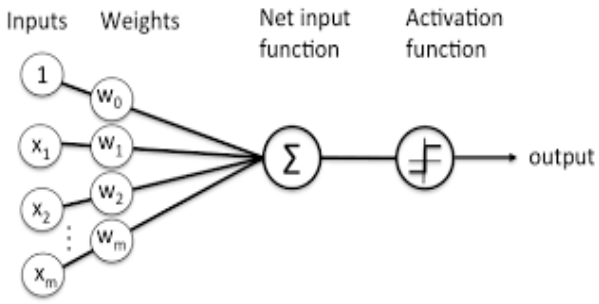

Figure 3. Neural Node

A neural network is a network of such functions, that may be defined as f(x)=f3(f2(f1(x))) and in the chain, f1 is the first layer, similarly f2 is the second layer and so on. A schematic representation of neural network is shown in Figure 4. Final layer is called the output layer and also while training the desired output of each layer is not visible so the middle layers are called the hidden layer. A Deep Neural Network (DNN) is a feed forward Artificial Neural Network (ANN), with multiple hidden layers and the higher level of abstraction.

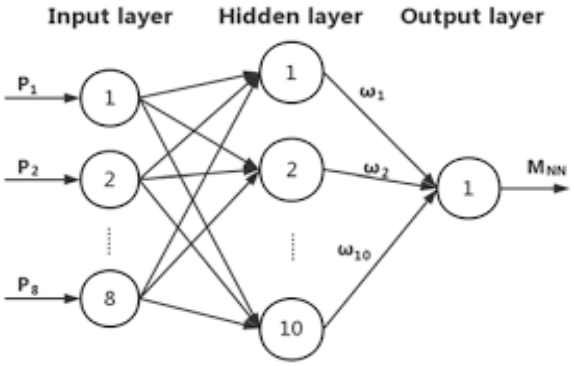

Figure 4. Simple Neural Network

In deep neural networks, learning requires minimizing the cost function, that in case of classification cost function is the difference between actual and the predicted label. Generally, gradient descent is used for this purpose. In modern neural network, use of Rectilinear Unit or relu is recommended as activation function. So limited data causes problem over fitting in DNN and randomly dropout some nodes from layers based on their probability "dropping out" that indicates removing units temporarily along with its incoming and outgoing edges. Is shown in figure 5.

Convolutional Neural Network

Convolutional Neural Network (CNN) is a class of deep, feed-forward artificial neural networks that has been successfully applied to analyzing visual imagery. CNN use variation of multilayer perceptron's designed to require minimal processing. CNNs has the ability to be able to detect abstract and complex features that makes them so attractive in image recognition problem.

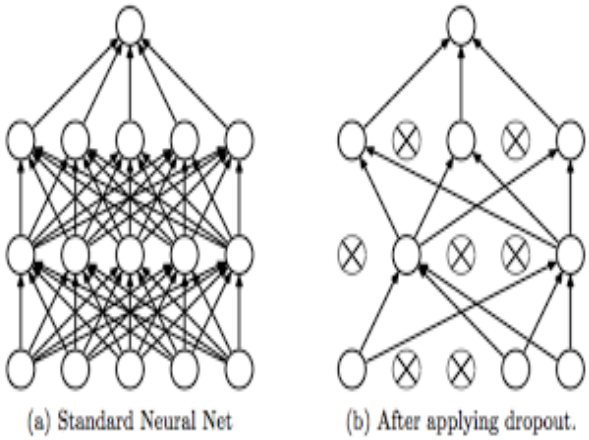

Figure 5.Droupout Neural Net

2D Convolutional Layer:

The most common type of convolution that is used is the 2D convolutional layer, and is usually abbreviated as conv2D. This layer creates a convolutional kernel that is convolved with the layer input to produce a tensor of outputs. 2D convolutional layer (e.g. spatial convolutional over images). A filter or a kernel in a conv2D layer has a height and a width. They are generally smaller than the input image and Conv2D filters extend through the three channels in an image (Red, Green and Blue). And each filter in this layer is randomly initialized to some distributions (Normal, Gaussian, etc.). So, by having different initialization criteria, then each filter gets trained slightly differently.

These are used in the first few convolutional layers of a CNN to extract simple features. Conv2D filters are used only in the initial layer of a CNN. They are put there to extract the high-level features from an image. The conv2D layer works fairly impressively.

Sequential Model Method:

Configures the model for training. There are two ways to build the Keras models: *sequential* and *functional*. The sequential API allows to create models layer-by-layer for most problems. The Sequential Model API is the way for developing deep learning models in most situations, but it has some limitations. For e.g., it is not straightforward to define models that may have: 1) multiple different input sources, 2) produce multiple outputs destinations, or 3) models that re-use layers. Is also a linear stack of layers.

MaxPooling2D:

Max Pooling operation for spatial data. This layer applies max pooling in two dimensions. In addition, *pooling* operations make up another important building block in CNNs. Then Pooling operation reduces the size of feature maps by using some function to summarize subregions, by taking the average or the maximum value.

Pooling works by sliding a window across the input and feeding the content of the window to a *pooing function*. This works much like a discrete convolution, but replaces the linear combination described by the kernel with some other functions.

Deep Learning:

It is a part of a broader family of machine learning methods that is based on learning data representations. Learning can be supervised, semi-supervised or unsupervised. The design of Deep learning for deep neural networks is applied to the fields that has pc vision, speech recognition, natural language processing, audio recognition and AI, drug design, medical image analysis, etc., that has produced the results comparable to human experts.

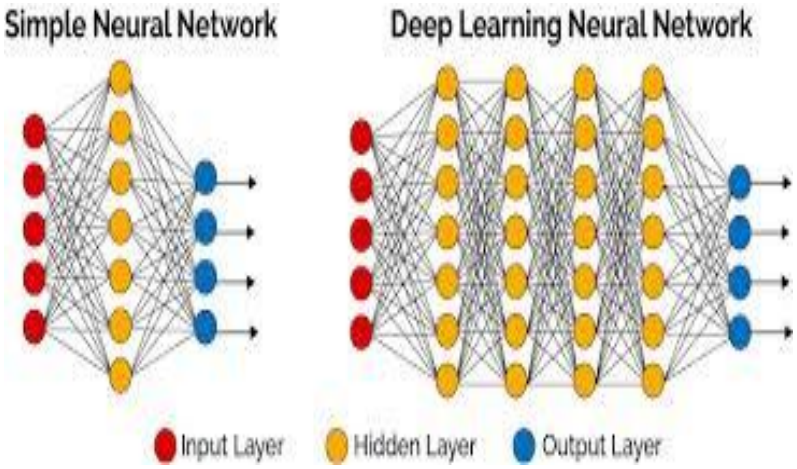

Figure 6. Deep Learning Neural Network

IV. RESULT

The basic working is described in figure 7. where the system captures an image of the sign language through webcam.

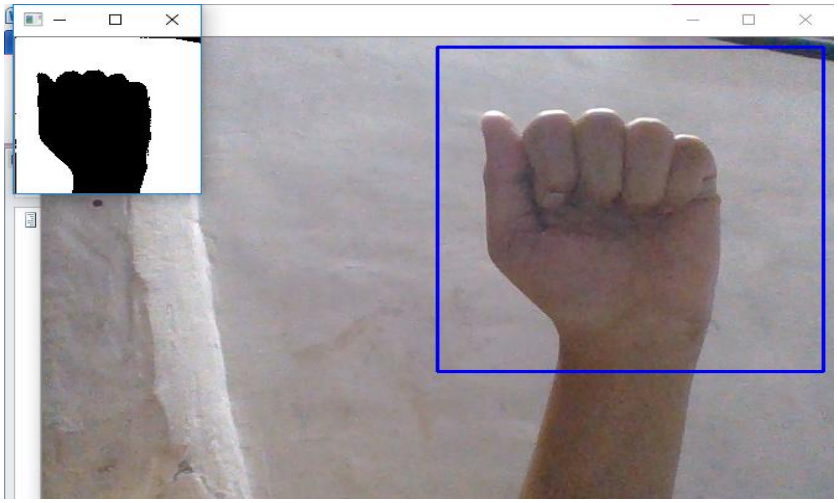

Figure 7. Window Capture an Image

The Next figure 8.Saves the captured image.

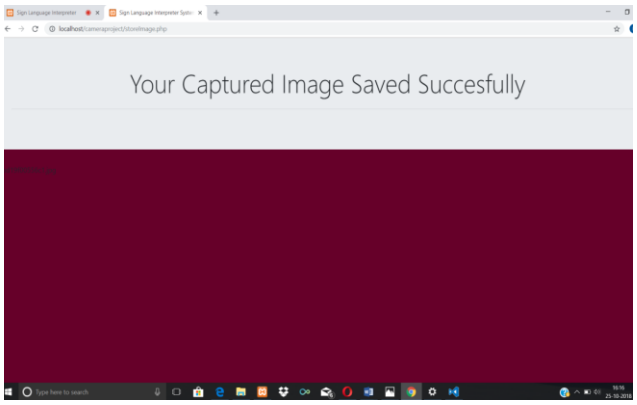

Figure 8. Saves the Captured Image

Following Figure 9.specifies the accuracy on the training model. We get 95% accuracy and 97% validation accuracy.

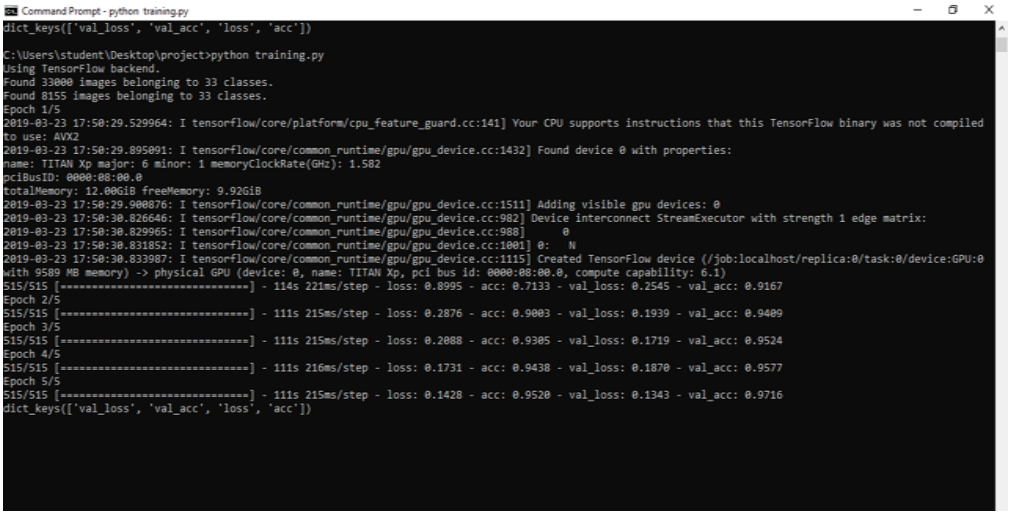

Figure 9. Training Code Output

Following figure 10.represents the accuracy graph of the training model.

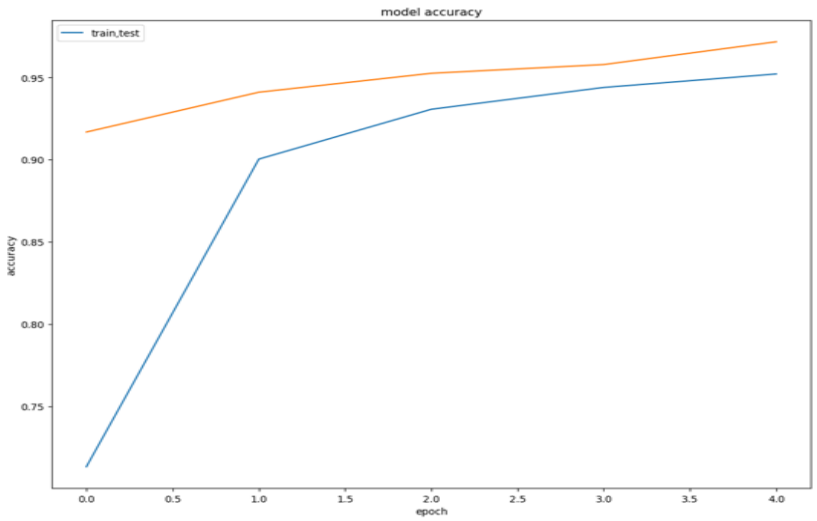

Figure 10. Accuracy graph of trained model

Following figure 11. Represents the loss graph of the training model

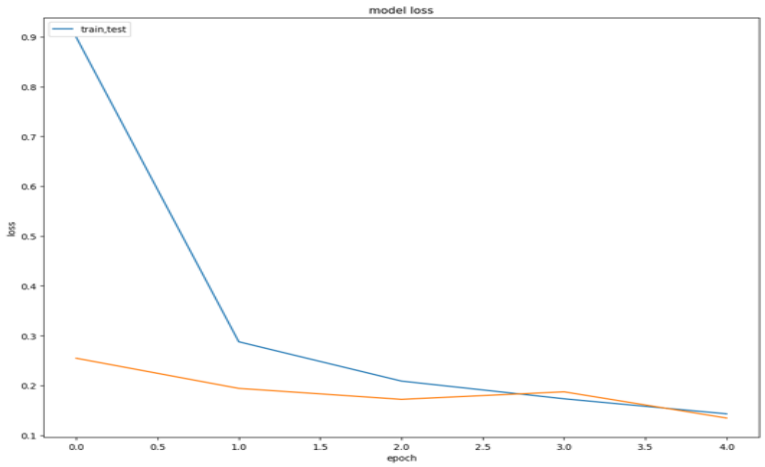

Figure 11. Loss graph of trained model

And figure 12. Shows the system prediction output for given sign input as Alphabet “A”.

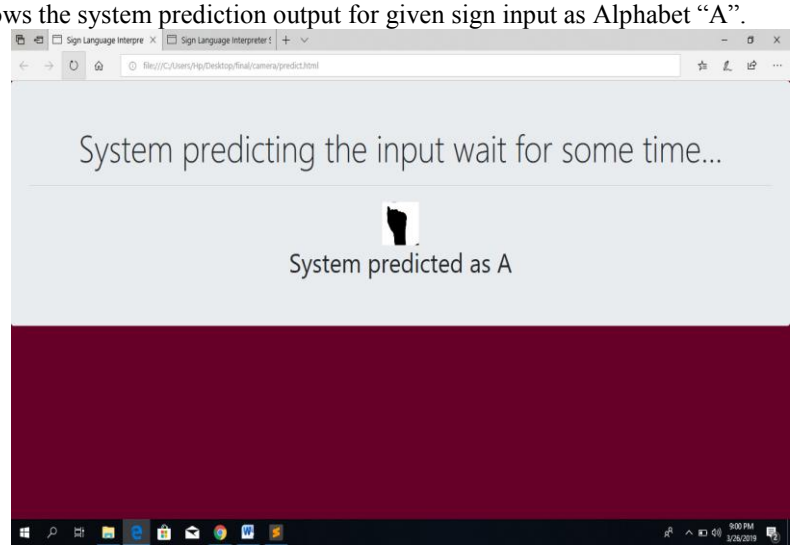

Figure 12. Window display prediction

V. FUTURE SCOPE

We are developing a project that would enable deaf people to get more involved in society. The idea of project is that, a camera-based sign language recognition system that would be in use for the deaf for converting sign language gesture to text and then speech. So, the objective is to design a solution that is intuitive and simple. Communication for majority of people will not be difficult.

This Sign Language Interpreter system will work as one of the futuristic of Artificial Intelligence and Computer Vision with user interface. It creates method to recognize hand gesture based on different parameters. And the main priority of this system is to be simple, easy and user friendly without making any special hardware. The further advanced version of the system might help the normal human being to convert their text into sign language which will build two-way communication. All computations will occur on single PC.

VI. CONCLUSION

Mute people are isolated from the most common forms of communication in today's society such as warning, or any other form of oral communication between people in regular daily activities. Sign language is a primary means of communication. So, to communicate using sign language there is a glove-based system through which communication is possible. But it has to be recalibrated every time whenever a new user uses a system. The connecting wires restrict the freedom of moment.

So, the solution to this problem is image processing with deep learning. The project is implemented in such a way that it does not require gloves. The gesture has to be formed in front of the camera and the output is given in the form of text or audio. Thus, we can conclude that the system can interpret American Sign Language in real time environment and can act as a communication device between a signer and a non-signer person.

ACKNOWLEDGMENT

We would like to forward our sincere thankfulness to the principal of our college for providing us with the chance to work on this paper. We would also like to thank him for providing us with all the resources which were required. We would also like to extend our thankfulness to the teaching and non-teaching staff as well as our colleagues for making this paper an enormous success and for spending their time and efforts in this paper.

REFERENCES

[1]. PratibhaPandey and Vinay Jain, "An Efficient Algorithm for Sign Language Recognition", International Journal of Computer Science and Information Technologies (IJCSIT), Volume 6 (6), 2015.

[2]. AkshayJadhav, GayatriTatkar, GauriHanwate and RutwikPatwardhan, "Sign Language Recognition", International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), Volume 7, Issue 3, March 2017.

[3]. Oyebade K. Oyedotun and Adnan Khashman, "Deep Learning in Vision-Based Static Hand Gesture Recognition", The Natural Computing Application Forum 2016.

[4]. Ashish S. Nikam and AartiAmbekar, "Sign Language Recognition using Image Based Hand Gesture Recognition Techniques", ResearchGate, [online], Available from:

[5]. *https://www.researchgate.net/publication/316732601_Sign_language_recognition_using_image_based_hand_gesture_recognition-techniques* (November 2016).

[6]. M. Mohandes, S. Aliyu and M. Deriche, "Prototype Arabic Sign Language Recognition using Multi-Sensor Data Fusion of Two Leap Motion Controllers", 12th International Multi-Conference on Systems, Signals & Devices, 2015.

[7]. CelalSavur and FeratSahin, "Real-Time American Sign Language Recognition System by using Surface EMG Signal", 14th International Conference on Machine Learning and Applications (ICMLA), 2015.

[8]. HeminaBhavsar and Dr. JeegarTrivedi, "Review on Classification Methods used in Image Based Sign Language Recognition System", International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), Volume: 5, Issue: 5, May 2017.

[9]. B. P. Pradeep Kumar and M. B. Manjunatha, "A Hybrid Gesture Recognition Method for American Sign Language", Indian Journal of Science and Technology, Volume 10(1), January 2017.

[10]. RabeetFatmi, SherifRashad, Ryan Integlia and Gabriel Hutchison, "American Sign Language Recognition using Hidden Markov Models and Wearable Motion Sensors", Transactions on Machine Learning and Data Mining, Vol. 10, No. 2 (2017) 41-55.

[11]. G. AnanthaRao, P. V. V. Kishore, A. S. C. S. Sastry and K. Shyamala, "Deep Convolutional Neural Networks for Sign Language Recognition", 2018.

[12]. Prof. NeelamPhadnis, RupeshPrajapati, VedantPandey, NupurJamindar and NeerajYadav, "Hand Gesture Recognition and Voice Conversion for Deaf and Dumb", Volume: 05 Issue: 04, April 2018.

[13]. LihongZheng, Bin Liang and Ailian Jiang, "Recent Advances of Deep Learning for Sign Language Recognition", 2017 IEEE.

[14]. Naresh Kumar, "Sign Language Recognition for Hearing Impaired People based on Hands Symbols Classification", International Conference on Computing, Communication and Automation (ICCCA), 2017 IEEE.
[15]. Jie Huang, Wengang Zhou, Houqiang Li and Weiping Li, "Sign Language Recognition using Real-Sense", 2015 IEEE.
[16]. Murat Taskiran, Mehmet Killioglu and NihanKahraman, "A Real-Time System for Recognition of American Sign Language By Using Deep Learning", 2018 IEEE.
[17]. *https://en.wikipedia.org/wiki/American_Sign_Language*, Accessed on 12/08/2018, at 02:30 p.m.
[18]. *https://www.kaggle.com/datamunge/sign-language-mnist*, Sign Language MNIST, Kaggle, 2017, Accessed on 28/08/2018, a 12:30 p.m.

Jessica Dias. "Sign Language Interpretation using Deep Learning." IOSR Journal of Engineering (IOSRJEN), vol. 09, no. 04, 2019, pp. 17-25.