An Interpreter for the Differently Abled Using Haptic Feedback and Machine Learning

Priya Prajapati[1], Gorrepati Sai Surya[2], M. Nithya[3]
Department of Electrical and Electronics Engineering,
Amrita School of Engineering, Bengaluru,
Amrita Vishwa Vidyapeetham, India.
[1]priyaprajesh15@gmail.com, [2]suryachowdari98@gmail.com, [3]m_nithya@blr.amrita.edu

***Abstract* — One of the most valuable blessings to humanity is the ability to see, hear, talk and respond according to the situation. However, some people are deprived of this. Such differently-abled people find it hard to communicate and hence they are permanently dependent on someone. To overcome this dependency and to bridge the communication barrier of such deprived, an interpreter can be used. This work proposes an Interpreter system which enables a two-way communication between the common people and the differently abled regardless of the kind of sensory disability the user has. The system implemented has two modules – an American Sign Language recognition module which uses machine learning algorithm to cater the people with vocal disability and a speech to haptic feedback conversion module which can be used for communicating to any kind of differently abled. Both these modules can be integrated together as a single compact device to address the communication barrier between different combinations of disabled people.**

Keywords—Raspberry Pi, haptic feedback, Python, Convolutional Neural Network , OpenCV, coin vibration motor, ASL recognition, machine learning.

I. Introduction

These days with the advancement of technologies, people are able to communicate effectively irrespective of the distance between them. Most people have no problem in communicating with each other but an exception is the hapless differently abled community such as the blind, deaf and dumb.

One of the oldest methods used by the visually impaired was the Braille script which enabled them to read the text that was embossed [1]. Another method that was used by the vocally disabled community to communicate was the use of gestures. Even though these methods were widely accepted by the differently abled community, they are neither very effective for two-way communication with every others nor are they suitable for people with other disabilities.

As technology advanced, many sensors and gesture recognition systems were introduced. Glove based systems used flex sensors which measure the change in resistance with the position of the finger, and then the value obtained is mapped to a particular gesture. These sensors have less tolerance value hence the gestures could be misinterpreted [2]. Another system was the braille embosser using servo motors which converted real time speech to braille, character by character. Six servo motors were used for the implementation of the device which made it less portable [3]. OMAR, a system which uses kinesthetic of the articulatory muscle for the reception of speech could make the disabled person understand the speech by moving and vibrating his fingers in one or two dimensions with the use of actuators [4]. Few systems use motion sensors like the inertial measurement unit (IMU) sensor, which is good at capturing hand orientations and hand-arm movements while the surface electromyography (sEMG) sensor distinguishes diverse hand shapes and finger movements when it is placed on the forearm. Along these lines, both have their own advantages in capturing distinctive data about a sign. In spite of the fact that the combination of these two integral modalities will improve the exhibition of the system, they aren't financially savvy [5]. To make the reception of speech simpler, haptic feedback devices like the coin vibration motor embedded in gloves were used, which encoded the speech character by character and then converted it to Morse code. The motors are very small thus they increased the portability of the system but Morse code has a lengthy encoding for each character hence it is a slow way to communicate [6].

Other well studied systems are the sign language recognition systems which uses a camera to capture images of the gestures, which is then classified using various methods. One of the first image classifiers converted the captured image to a binary image to find the region of active fingers by using the hit or miss technique to locate transitions, after which the finger count is found to map it to a gesture. This technique could classify very few gestures as most of the gestures had similar patterns [7]. Edge detection also became very commonly used method for classification, hence for the gesture recognition, the fingers are given an up or down position depending on the Euclidean distance from the reference and then classified [8]. Another method required the vocally disabled person to wear colored rings, so that the erect fingers could be identified with their distinct rings which is then recognized and classified [9]. With the advancement of software technologies, machine learning came into picture. An OCR(Optical Character Recognition) text recognition uses a camera to capture the text which is then converted to digital form, and incorporates text to speech conversion which is helpful for people with visual impairment. But this requires the person to always carry something to write on, consumes time and is also tedious for long conversations [10]. In some cases, Principle Component Analysis (PCA) is used for extracting the best fitting region which can be used with fuzzy-c-means clustering or K Nearest Neighbor (KNN) for classification of the gestures [11-12].

A detailed analysis of various techniques used for the communication to the differently abled shows that most of the systems have used Sign Language Recognition and flex sensors. The image recognition system used were either too advanced that needs high end processors which made them not portable and costly or not at all advanced and cannot tolerate little changes in the hand gesture. The aim of this work is to build an interpreter that incorporates the desirable features of the above literatures in addition to addressing the limitations they have. Coin vibration motors which incorporate the 5x5 tap encoding for characters is used for haptic feedback. A machine learning Convolutional Neural Network (CNN) model is used for American Sign Language

(ASL) gesture classification [13-16, 17]. The design and implementation details of the interpreter to enable two-way communication between the normal and differently abled are discussed in detail in the following sections.

This paper is organized as follows. Section II projects the model outline of the interpreter with all the peripherals coupled to Raspberry Pi. The two modules of the Interpreter along with their block diagram and implementation details are also discussed. Section III discusses the results of the two modules and finally section IV lays out the conclusion and future scope of this work.

II. SYSTEM CONSTRUCTION AND IMPLEMENTATION

A. *System Diagram*

Fig. 1 displays the complete outline of the system with Raspberry Pi and all the peripherals that are needed to make the two-way interpreter. The peripherals include a Pi camera, USB microphone, a display, coin vibration motors and speakers.

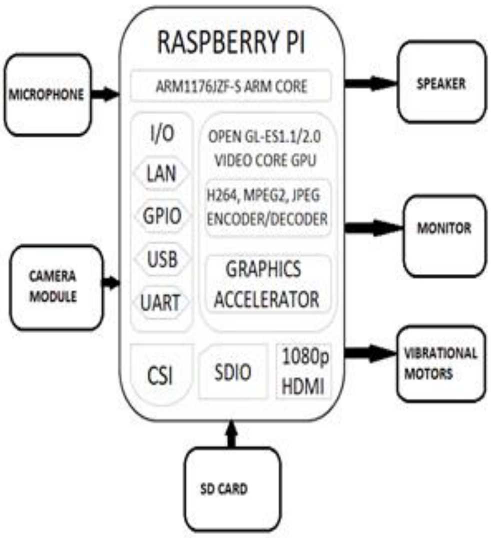

Fig. 1. System outline

The proposed interpreter can be seen as two modules. Module 1 is used for the communication from a non-disabled person to a differently abled person, whereas module 2 deals with the converse communication which are explained in detail in the sub sections B and C. The specifications of the components required in this system are tabulated in Table I.

TABLE I. SPECIFICATIONS OF COMPONENTS USED IN THE INTERPRETER

Component	Specifications	Quantity
Vibration motor	1-4V, 66mA, 12000rpm	10
Pi camera	5MP, CSI	1
USB microphone	22x18x7mm	1
Raspberry Pi	Model 3B	1
LCD Display	Use in 4-bit mode	1
SD Card	32GB	1

B. *Communication from Normal Person to the Differently Abled Person (Module 1)*

This module explains about the communication from a non-disabled person to a differently abled person. For most people, the easiest way of communication is talking, and if the person at the receiving end suffers from hearing impairment, then other means of communication are required. The different stages involved in this communication are shown in Fig. 2.

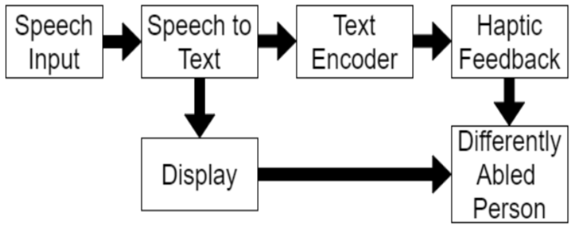

Fig. 2. Block diagram of Module 1

In the above block diagram for communication, the input to this module is speech, which is converted to text, and then displayed on an LCD for the non-visually impaired. The text output at the second stage is also encoded character by character, which then mapped to vibrations (haptic feedback) and the same can be felt by the disabled person on his/her fingers. The Interpreter module1 proposed in this work addresses communication between various combinations of disabilities. TableII. lists possible combinations of impairment and the corresponding method through which the disabled can receive information from the speaker side.

TABLE II. MODULE 1 COMMUNICATION COMBINATIONS

From	To (type of disability)	Method used to convey
Person with no disabilities	Visual disability	Speech
	Auditory disability	Text(display)
	Vocal disability	Speech
	Sight and auditory disability	Haptic feedback
	Sight and vocal disability	Speech
	Auditory and vocal disability	Text(display)
	Sight, auditory and vocal disability	Haptic feedback

The module 1 uses Raspberry Pi 3B as the controller which also offers a common interface for all the hardware and software components used. A USB microphone is used for high-quality speech recording. This is then converted to text using suitable python libraries and code. Coin vibrational motors are used to produce vibrations corresponding to the text generated. These coin motors are attached to the gloves that have to be worn by the visually impaired person. The vibrations are generated using 5x5 tap method for alphabets and the same is done for numbers. The vibrations encoding for alphabets and numbers are given in Table III and Table IV. The tables shown describes which finger should vibrate and the number of vibrations generated. The assumption made here is that the user is thorough with tap encoding.

TABLE III. ALPHABET ENCODING

Finger Number →

Number of Vibrations ↓

	1	2	3	4	5
1	A	B	C/K	D	E
2	F	G	H	I	J
3	L	M	N	O	P
4	Q	R	S	T	U
5	V	W	X	Y	Z

TABLE IV. NUMBER ENCODING

Number of Vibrations ↓ / Finger Number →	1	2	3	4	5
1	0	1	2	3	4
2	5	6	7	8	9

C. *Communication from the Differently Abled Person to Normal Person (Module 2)*

Module 2 explains the communication techniques used by the differently abled people to communicate with other people. This module comes handy only for the people with vocal disabilities, as other people can make use of speech directly. The step by step process involved in this module is shown in Fig. 3.

Fig. 3. Block diagram of Module 2

Here, the person who is familiar with the ASL uses hand gestures which are captured and processed. The prediction takes place and the sentence is generated. After this, the sentence is converted to speech. Table V. lists possible combinations of impairment and the corresponding method through which the disabled can convey information to the listener.

TABLE V. MODULE 2 COMMUNICATION COMBINATIONS

From (type of disability)	To	Method used to convey
Sight disability	Person with no disabilities	Speech
Auditory disability		Speech
Vocal disability		Hand gesture
Sight and auditory disability		Speech
Sight and vocal disability		Hand gesture
Auditory and vocal disability		Hand gesture
Sight, auditory and vocal disability		Hand gesture

The first step for the image recognition is to create a huge dataset with large number of images for each of the 29 classes. These 29 classes consist of 26 alphabets and 3 special characters. The data set of the images should have variations such as position of the hand, spacing, brightness and contrast. Also, it should be made sure that different hands are used during creation of the dataset. In this work, 4000 images are created for each class. The next step is to augment the dataset (preprocessing) in order to increase variations in the images, to enlarge the dataset. The augmentation techniques implemented are shearing, rotation to a small range, width and height shifts.

The modified dataset is split into two parts, as training data (90%) and validation data (10%). The processes involved are shown in Fig. 4.

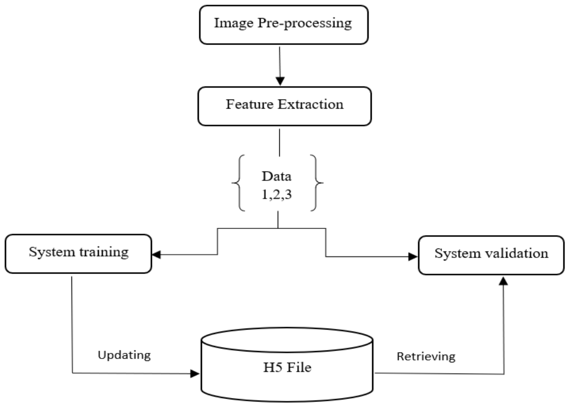

Fig. 4. Algorithm for CNN model creation

A sequential CNN model is initialized and the model is trained using the augmented data. The optimizer used for the model is an adaptive movement estimator (ADAM) which gives a better accuracy in less time compared to other optimizers available for image classification [18]. The algorithm validates itself by taking an image from the validation dataset and checks for correctness of prediction. The above algorithm runs until good accuracy is achieved. The training process is stopped when losses are very less. The training should be stopped before overfitting. To get a better accuracy, dropout is used to slow the training process.

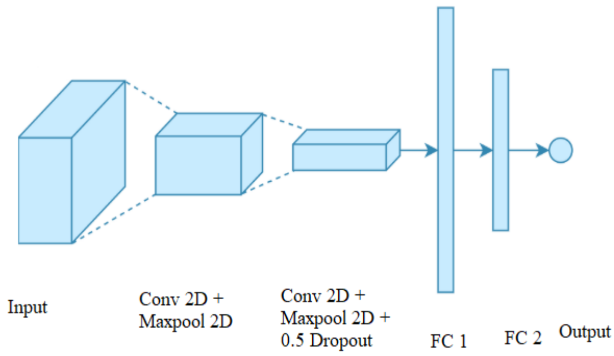

Fig. 5. CNN architecture

The CNN model used has two convolutional and pooling layers with a dropout followed by fully connected layers to get the output of image as shown in Fig. 5. The model is saved when the validation loss is minimum, because further training would result in overfitting of the trained data. The trained model weights along with the CNN architecture is saved in a H5 file to avoid training every time a prediction has to be made. This model is then uploaded into a prediction code which uses OpenCV for real time image capturing. The image captured is converted to a numpy array which is a grayscale image. This array is sent to the pre trained model where it is classified and the corresponding alphabets/numbers are printed on the screen. This process continues until the user manually stops the prediction. The complete sentence formed is then extracted and read out using the text to speech module.

III. RESULTS AND DISCUSSION

A. *Module1- Normal Person to the Differently Abled Person Communication*

Speech is recorded using the USB microphone on Raspberry Pi, the speech is then converted to text and displayed on the LCD screen. The text is converted to haptic feedback by the 5x5 tap method; these vibrations have sufficient delay between them so that the user recognizes each vibration correctly. As the number of GPIO pins on the Raspberry Pi are limited, a de-multiplexer 74LS138 is used to reduce the number of GPIO pins required in the circuit. Three GPIO pins of Raspberry Pi are connected to select lines of de-multiplexer which gives eight outputs, from which five are connected to the coin vibration motors. Similar connections are made for the second de-multiplexer. A total of ten motors are used, five of which is for alphabet encoding and the remaining motors for number encoding.

The simulation model of module 1 is shown in Fig. 6, with one de-multiplexer and with eight LED's which represent the coin vibrational motors in hardware circuit. The select lines of the de-multiplexer are controlled by the switches B1, B2 and B3. The enable pins E2, E3 should be active low and E1 should be active high. Each combination of select lines turns on one specific LED which can be seen as vibration of the coin motors.

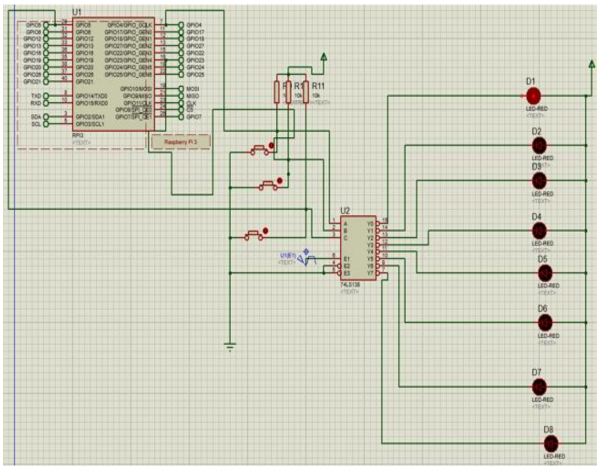

Fig. 6. Simulation circuit of module1

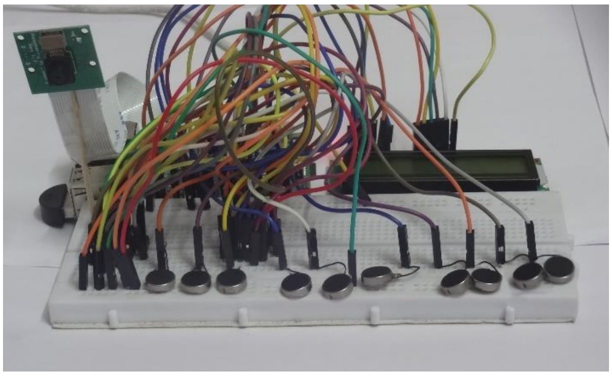

Fig. 7. Hardware implementation of module1 using Raspberry Pi

Once the outputs are verified in the simulation, hardware circuit is rigged up as shown in Fig. 7. The code is uploaded on the Raspberry Pi. After the input speech is converted to a text string, corresponding motors are vibrated by enabling the select lines of the de-multiplexer according to the character and number encoding. Thus communication from a non-disabled to a disabled person is achieved using haptic feedback.

B. *Module 2- Differently Abled Person to Normal Person Communication*

The communication from the disabled person requires a prediction methodology for which a dataset consisting of 29 classes is created which is shown in Fig. 8. After performing data augmentation, the dataset is sent to the model for training.

Fig. 8. Sample dataset

The model is trained initially for 100 epochs without the use of dropouts. This leads to overfitting of the training data which is depicted in the performance curves shown in Fig. 9 and Fig. 10.

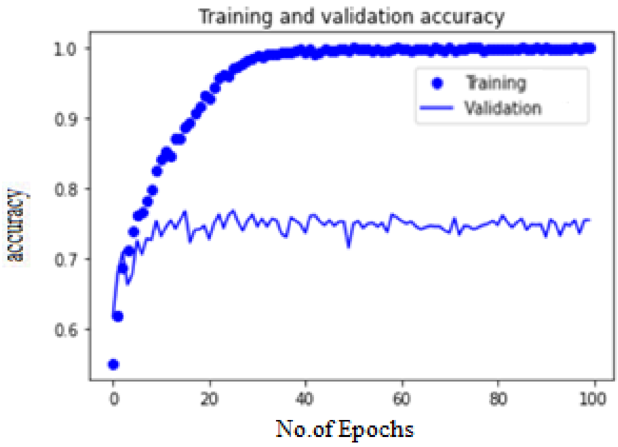

Fig. 9. Training and validation accuracy vs epoch

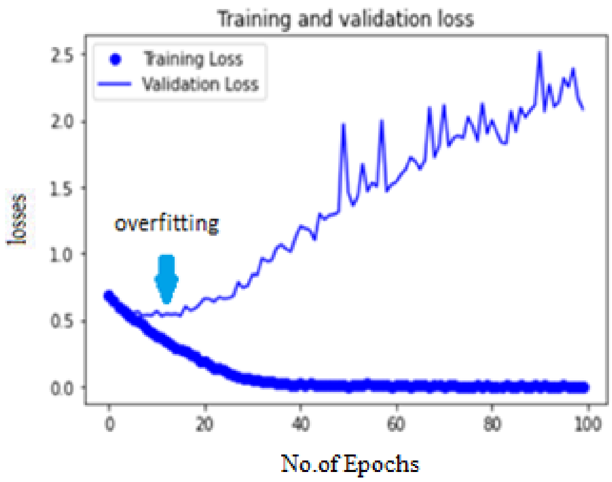

Fig. 10. Training and validation losses vs epoch

After the use of dropouts, it is observed that the training and validation accuracies has stabilized around 90 epochs which is shown in Fig. 11.

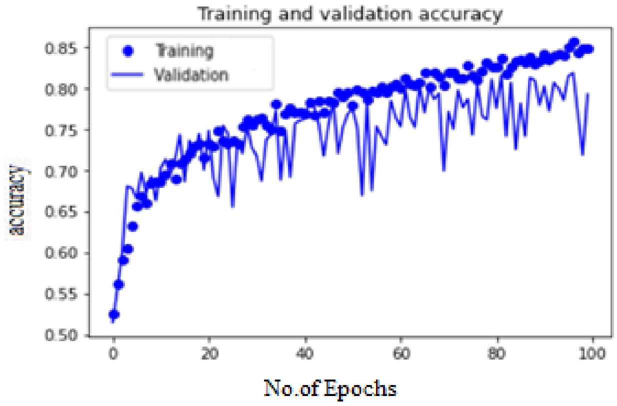

Fig. 11. Training and validation accuracy vs epoch with dropout

It can also be observed from Fig. 12 that the training and validation losses have reduced exponentially during the 100 epochs.

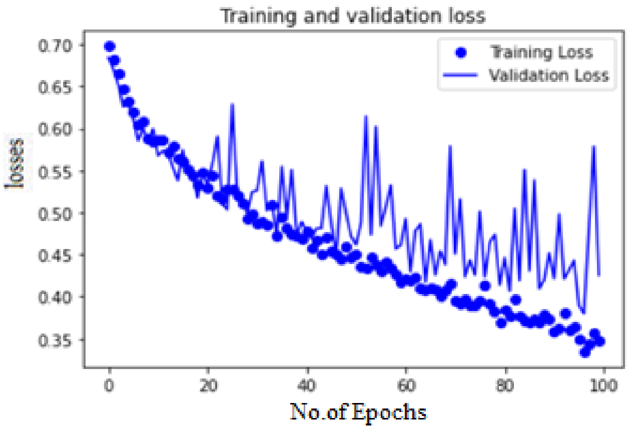

Fig. 12. Training and validation loss vs epoch with dropout

The training process was tested with various changes in parameters like optimizer used, augmentation ranges, dropout percentages, etc. Finally the most optimum result was found out with the use of ADAM optimizer and 50% dropout. The values of augmentation attributes used for this training are rotation = 5, width shift = 0.1, height shift = 0.1, shear range = 0.2 with mode as nearest. The testing parameters are tabulated in Table VI.

TABLE VI. ACCURACY AND LOSS VALUES

No. of Epochs	Loss (in %)	Accuracy (in %)	Validation Loss (in %)	Validation Accuracy (in %)
10	60.88	65.85	56.18	70.1
20	55.15	71.95	57.96	69.3
30	51.77	74.25	49.13	74.9
40	50.83	74.90	45.79	77.6
50	46.35	78.75	47.23	77.8
60	45.96	77.10	47.08	77.9
70	43.47	80.05	44.19	78.4
80	42.27	80.85	50.62	75.8
90	40.93	81.05	42.91	81.8
100	38.37	83.35	40.15	82.3

The predicted text can be displayed in the screen as shown in Fig. 13.

Fig. 13. Predicted text output for the corresponding ASL

IV. CONCLUSIONS AND FUTURE WORK

This paper discusses various methods for communication between the differently abled and others. The Interpreter developed in this work is a low-cost two-way interpreter which provides effective and efficient communication between the two communities. In module 1, the speech is converted to text and then the alphabets are encoded as vibrations or displayed as text. This was found to be better than the OCR (Optical Character Recognition) as the drawback of the need to carry something to write is eliminated in this work. The other method Braille Embosser had the limitation that it could only be used by the blind people whereas the Interpreter proposed here can be used by people with all the disabilities as discussed. In the second module, the person knowing the American Sign Language, uses hand gestures in front of a camera, and a real time hand gesture recognition is carried out using a pre-trained CNN model. In other similar works, flex sensors with little tolerance were used, which did not give accurate results. This interpreter not only gives a better accuracy but it also eliminates the need of sensors. Combining both the modules proposed will enable communication between two differently abled people having diverse inability. For example, the gestures can be converted to haptic feedback thereby increasing the applications.

REFERENCES

[1] S.Padmavathi, Manojna K.S , Sphoorthy Reddy .S and Meenakshy.D, "Conversion of braille to text in english, hindi and tamil languages," *International Journal of Computer Science, Engineering and Applications (IJCSEA)*, June 2013, Vol.3, No.3, pp. 19-32.

[2] Sunita V. Matiwade, Dr.M.R.Dixit, "Electronic Device for Deaf and Dumb to Interpret Sign Language of Communication," *International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801, ISSN(Print): 2320-9798*, Nov 2016, Vol.4, Issue.11, pp.19662-19669.

[3] M.S.Ghute, S.Soitkar and K.P.Kamble, "A Novel Approach to Communicate with Deaf Dumb and Blind Person," *International Journal of Engineering Research & Science (IJOER)*, April 2017, Vol.3, Issue.4, pp. 23-26.

[4] Eberhardt SP, Bernstein LE, Coulter DC, Hunckler LA, "OMAR a haptic display for speech perception by deaf and deaf-blind individuals," *In Proceedings of IEEE Virtual Reality Annual International Symposium*, Sept 1993, pp. 195-201.

[5] Lu Sun, Roozbeh Jafari and Juan Wu, "A Wearable System for Recognizing American Sign Language in Real-Time Using IMU and Surface EMG Sensors," *IEEE Journal of biomedical and health informatics*, Sept 2016, Vol. 20, No. 5.pp. 1281-1290.

[6] B.Swaminathan, J.Lokesh Kanna, B.Navya, Nakul Gopinath, "GLOVE for Speech Recognition," *International Journal of Computer Science and Mobile Computing(IJCSMC)*, May 2020, Vol.9, Issue.5, pp. 90–99.

[7] P Raghu Veera Chowdary, M Nagendra Babu, Thadigotla Venkata Subbareddy, Bommepalli MadhavaReddy, V Elamaran, "Image Processing Algorithms for Gesture Recognition using MATLAB," *IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT),* Ramanathapuram, India, May 2014, pp. 1511-1514.

[8] P.Subha Rajam, Dr.G. Balakrishnan, "Real time Indian sign language recognition system to aid deaf-dumb people," *IEEE 13th international conference on communication technology*, Jinan, China, Sep 2011, pp. 737-742.

[9] Koli P.B., Chaudhari Ashwini, Malkar Sonam, Pawale Kavita and Tayde Amrapali, "Image Processing Based Language Converter for Deaf and Dumb People," *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), e-ISSN: 2278-2834, ISSN: 2278-8735,* 2015, pp. 25-30.

[10] Anish Kumar, Rakesh Raushan, Saurabh Aditya, Vishal Kumar Jaiswal, Mrs. Divyashree Y.V, "An Innovative Communication System For Deaf, Dumb and Blind People," *International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN:2321-9653,* June 2017, Vol.5, Issue.6, pp. 42-52.

[11] M. Ashraful amin, Hong yan, "Sign language finger alphabet recognition from GABOR-PCA representation of hand gesture," *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong*, China, 2011, No.4, pp. 245-259.

[12] Sruthi Upendran, Thamizharasi. A," American Sign Language Interpreter System for Deaf and Dumb Individuals," *International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT),* Kanyakumari, India, Jul 2014, pp. 1477-1481.

[13] Saad AL-ZAWI, Saad albawi, Tareq Abed MOHAMMED," Understanding of a Convolutional Neural Network," I*nternational Conference on Engineering and Technology (ICET),* Antalya, Turkey, Aug 2017, pp. 1-6.

[14] Geoffrey E. Hinton, Ilya Sutskever, Alex Krizhevsky, "ImageNet Classification with Deep Convolutional Neural Networks," *In Advances in neural information processing systems,* 2012, pp. 1097-1105.

[15] R Nandakumar, Arjun A M, Sreehari S, "The interplay of hand gestures and facial expressions in conveying emotions a CNN–based approach," *Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC)* IEEE Xplore Part Number: CFP20K25-ART, Erode, India, 2020, pp. 833-837.

[16] Rajalakshmi Prithvi raj, Rahul Ravi, Yadhukrishna S.V, "A Face Expression Recognition Using CNN & LBP," *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE Xplore Part Number:CFP20K25-ART, ISBN:978-1-7281-4889-2, Erode, India, 2020, pp. 684-689.

[17] Shivashankara S, Srinath S, "American Sign Language Recognition System: An Optimal Approach," *International Journal of Image, Graphics & Signal Processing,* Aug 2018, Vol.10, Issue.8, pp. 18-30.

[18] K.Thammireddy,V.ValliKumari, S.V.G.Reddy, "Optimization of Deep Learning using various Optimizers, Loss functions and Dropout," *International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878,* Dec 2018, Vol.7, Issue.4S2, pp. 448-455.