RESEARCH ARTICLE | MAY 22 2023

S2S translator using CNN

Jeethu Philip ✉; Y. Lakshmi Prasanna; G. Kiran Kumar; P. Neelakanta Rao

AIP Conference Proceedings 2492, 020014 (2023)
https://doi.org/10.1063/5.0113182

S2S Translator Using CNN

Jeethu Philip [1,a)], Y.Lakshmi Prasanna[2,b)], G.Kiran Kumar[3,c)] , P.Neelakanta Rao[1,d)]

[1]*Department of Information Technology, MLR Institute of Technology, Hyderabad, India.*
[3]*Department of Computer Science Engineering, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India.*
[3]*Department of Computer Science Engineering, Chaitanya Bharathi Institute of Technology, Hyderabad, India.*

[a)] Corresponding Author: jeethusamson@gmail.com
[b)]prasanna.yeluri@gmail.com
[c)] ganipalli.kiran@gmail.com
[d)]neelakantharaop@gmail.com

Abstract. Sign language is a language used by dumb and deaf people to communicate with normal people. Normal people use sounds, unlike them, this language uses visual communication to convey the thoughts of dumb and deaf people. Sign language is achieved by continuously showing hands, the orientation of fingers, and facial expressions. In this project, we will develop a programmatic model that converts voice to sign language and also sign language to voice/text. we may be using different APIs (python modules or Google API) and natural language processing semantics to break the text into a large number of smaller understandable words which require machine learning as a part. Predefined alphabet signs are given as inputs to the model. So this can use Artificial Intelligence technology to translate audio into sign language and sign language to text.

Keywords: Artificial Intelligence, Natural Language Processing, Communication, Orientation, Predefined. API

INTRODUCTION

Sign language is a language used by dumb and deaf people for communication. Normal people use sounds, unlike them, this language uses visual communication to convey the thoughts of dumb and deaf people. Sign language is achieved by continuously showing hands, the orientation of fingers, and facial expressions. Each country has its own sign language. Britishers have British Sign Language (BSL) and Americans have ASL (American Sign Language). Both these languages are different and American Sign Language using people could not understand British Sign Language and vice-versa. Some countries adopt options of sign language in their sign languages. No one invented sign language. No one knows the exact birth of sign language. Some sources suggest sign Language was developed nearly 200 years ago by combining local sign languages and LSF (French Sign Language). Today's sign language includes some elements of French Sign Language and the original local sign languages. Over time, today's sign language evolved as a mature language. Sign Language is a separate language and distinct from the English language [1-6]. They still contain some similar signs; they'll now not be understood by every other's user. Sign Language may be a language fully separate and distinct from English. Sign language has all the options of the language, like pronunciation, word formation, and ordination. Every language has different ways of signaling different functions, such as asking a help rather than making an order, languages differ in these cases. English speakers raise their voices while asking a question and adjusting order to the words. Sign Language using people to ask questions with their eyebrows by raising them, widening eyes, and tilting bodies forward direction [7-12]. In different languages, specific ways in which of expressing ideas in sign communication vary. In addition to individual variations in expression, signing has regional accents and dialects. just as certain English words are spoken differently in different parts of the country, sign language has regional variations in the rhythm of signing,

Advancements in Aeromechanical Materials for Manufacturing: ICAAMM-2021
AIP Conf. Proc. 2492, 020014-1–020014-9; https://doi.org/10.1063/5.0113182
Published by AIP Publishing. 978-0-7354-4438-6/$30.00

pronunciation, slang, and signs used. Other sociological factors, including age and gender, can affect sign language usage and contribute to its variety, just as with spoken languages. Sign language contains figure-spelling and English letters are shown using fingers. In the finger-spelled alphabet, each letter has distinct hand shape and figure orientation. To be able to translate in the way that we would like to do i.e. either text or speech to sign language or sign language to text or speech. To be able to give input in terms of text or speech or as a live webcam feed. Must be able to give continuous input to the model for it to classify. The model must be able to classify with at least 60% accuracy. The output should be displayed with enough clarity to understand [2]

LITERATURE SURVEY

"Deaf Mute Communication Interpreter"-[1]: This paper aims to cover the various prevailing methods of deaf-mute communication interpreter system. The two broad classification of the communication methodologies used by the deaf –mute people are - Wearable Communication Device and Online Learning System. Under Wearable communication method, there are Glove based system, Keypad method and Handi com Touch-screen. All the above mentioned three sub-divided methods make use of various sensors, accelerometer, a suitable micro-controller, a text to speech conversion module, a keypad and a touch-screen. The need for an external device to interpret the message between a deaf –mute and non-deaf-mute people can be overcome by the second method i.e. online learning system. The Online Learning System has different methods. The five subdivided methods are- SLIM module, TESSA, Wi-See Technology, SWI_PELE System and Web-Sign Technology.

Hand gesture recognition and voice conversion system for dumb people [2]: Authors presented the static hand gesture recognition system using digital image processing. For hand gesture feature vector SIFT algorithm is used. The SIFT features have been computed at the edges which are invariant to scaling, rotation, addition of noise.

Hand Gesture Recognition for Sign Language Recognition: A Review in [3]: Authors presented various method of hand gesture and sign language recognition proposed in the past by various researchers. For deaf and dumb people, Sign language is the only way of communication. With the help of sign language, these physical impaired people express their emotions and thoughts to other person.

A Review on Feature Extraction for Indian and American Sign Language in [9]: Paper presented the recent research and development of sign language based on manual communication and body language.

Sign language recognition system typically elaborate three steps pre processing, feature extraction and classification. Classification methods used for recognition are Neural Network (NN), Support Vector Machine (SVM), Hidden Markov Models (HMM), Scale Invariant Feature Transform (SIFT),etc.

Speech and speaker recognition system using artificial neural networks and hidden Markov model in [11]: Paper isAiming towards automatic machine learning by human, a methodology for speech recognition with speaker identification based on Hidden Markov Model for security is a demand of science. This methodology to identify speaker and detection of speech. Here the acquisition of speech signal, analysis of spectrogram, neutralization, extraction of features for recognition, mapping of speech using Artificial Neural networks is presented. [3]

PROPOSED SYSTEM

In this project, we will use image processing and speech recognition, and neural networks to create an accurate model which would classify the input in whichever way necessary. Predefined data sets of sign language will be used for training our neural network model for classification. Image Processing will be used to take the images as input and for making them suitable for classification. Similarly, Speech Recognition will be used to convert audio to text and then classify it to the respective sign language image correspondingly.

Neural Network

A neural network is composed nodes. A neural network is an artificial neural network, for solving Artificial Intelligence problems. The biological neurons are connected and connections contains weights. A positive number weight is a powerful connection, while a negative number weight means less powerful connections [13&14]. All the

input weights are modified and summed. This is called a linear combination. The amplitude of the output is controlled by the activation function. The output between 0 and 1, or −1 and 1 are acceptable. These artificial neural networks are also used for prediction, adaptive management, and applications wherever they will be trained with a dataset. Self-learning will occur inside the networks.[4]

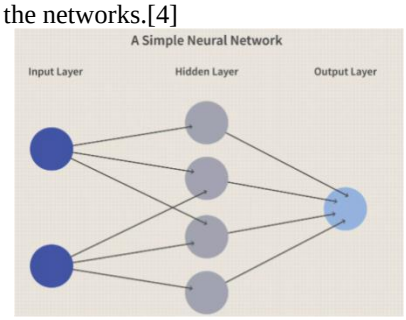
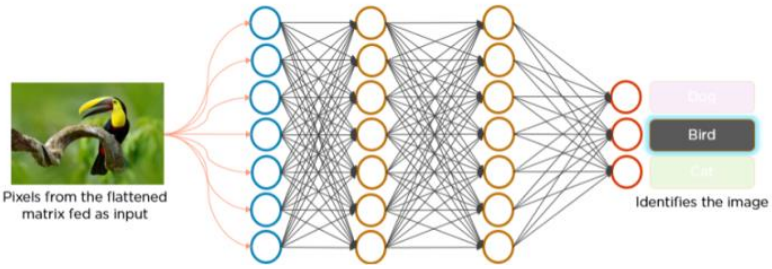


FIGURE 1 Neural Network

CNN

- A CNN or convolutional neural network or ConvNet is a deep learning algorithm that takes an image as input and assigns importance to those images, so it can differentiate one from the other. The pre-processing in a CNN is very lower as compared with other classification algorithm[5]



.

FIGURE 2 CNN reading the image of bird

- **FIGURE 2**, explains that camera detects the hand posters and orientation of fingers to identify the letters. The below figure shows the fingers of human hand.

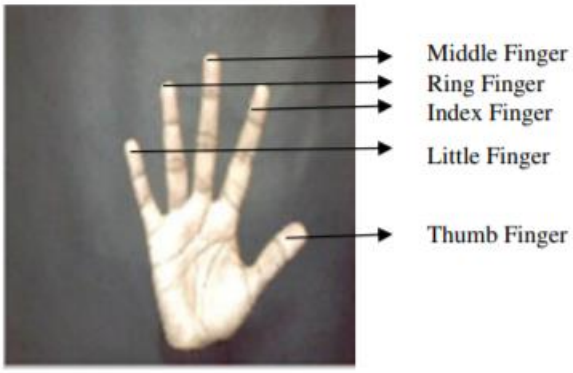


FIGURE 3 Name of fingers in hand

- The below **Figure 4** shows how the system process the image of the hand when you place in front of the camera. The difference between the original image, segmented image, and Edge detected image is clearly shown below.[6]

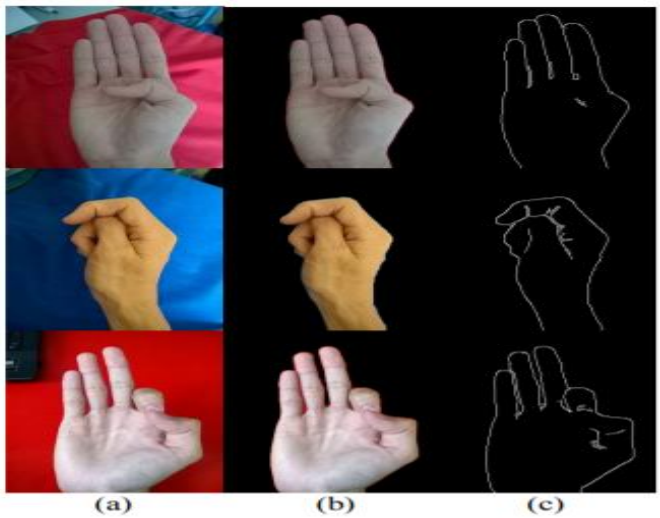


FIGURE 4. sign languages (a) original Image (b) segmented Image (c) Edge detected Image

- The below figure shows the signs of alphabets in a English language. We trained the model with all the 26 alphabets. When the signer shows his hand to the camera it will identify the letter that hand is showing and show the letter on the console.[7]

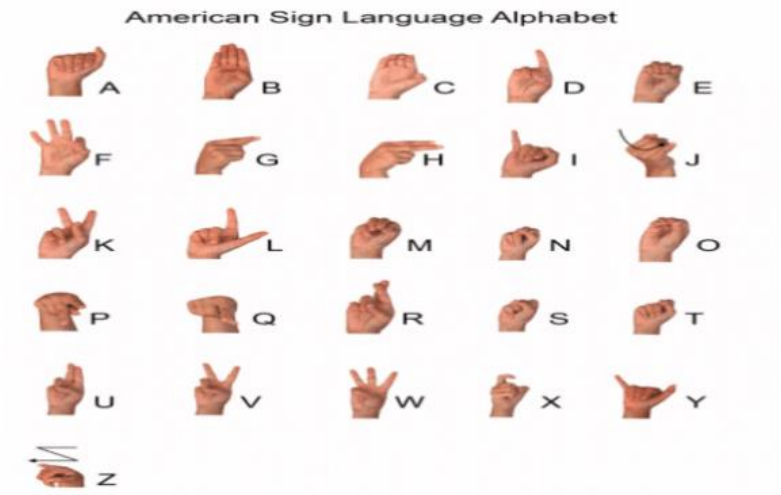


FIGURE 5 American Sign Language Alphabets

IMPLEMENTATION

Getting the Dataset

Datasets of predefined sign language were taken from Kaggle. It had approximately 10,000 images corresponding to each alphabet including both training and test data sets. An image corresponding to each alphabet was taken and displayed using matplotlib so that users can see what the corresponding sign to each alphabet was. Datasets of some frequent sign language gestures were also gathered and were used similarly as mentioned above.

Pre-Processing the Dataset

The training data of the sign images were taken and was cleaned to make them ready for classification. Each image was converted to 64X64 and the color was changed to monochrome. All the images were appended to an image array. The image array was then converted to a NumPy array. The array was normalized by converting the values to 'float32' and dividing it with the total sum of the pixel values i.e. 255. The labels corresponding to each

image were stored in an array as well and were converted to categorical values using the internal function present in the Keras module. The whole image data was split into training and testing using the internal function present in the scikit-learn module. The shape of the train and test data were printed and returned back for the model to consume.[8]

Creating Model

After pre-processing of the training data, we will create a CNN model which is used for classification. Keras module is used extensively here to create the model. We created a sequential model with 16 nodes in the first layer having the kernel shape [3,3] and the input shape was [64,64,3]. The activation function used was "Relu". The next layer had 32 nodes having the kernel shape [3,3] with a similar activation function. We added a MaxPool layer of two-dimension before getting the output to the next layers. The next layer connected to 32 nodes which were further connected to 64 nodes having a MaxPool2D layer. The output of these 64 nodes was fed to the next layer which had 128 nodes connected further to 256 nodes. After the MaxPool2D layer, batch normalization was done here. We flattened the output of the layers here and set the dropout to "0.5". The last layer of processing had 512 nodes where also kernel_ regulizer was set to "0.001". At last, we had the output layer of 29 nodes since we have 29 labels or classes. SoftMax function was used to get the output in terms of integers for easy processing and accurate classification. The model was compiled with an "adam" optimizer and loss and accuracy metrics were recorded while fitting the data into the model. Summary of the model was printed in case anyone in the future will be able to use it and the model was saved for further use.

Training the CNN Model

After the model is created, we need to fit our train data into it so that the model will learn about the data through training and then it would be used for further classifying real-time data. We use the "fit" method to give our training data. The batch size is set as 64 which means that images in will be fed to the CNN model in batches with 64 each in one batch. The epoch is set to be 5 i.e. for every five minutes the data will be fed to the model. The validation split is set as "0.5" which means that the data set is divided into two parts, one for training and other for validation. After the training is done, an object is returned which has all the metrics that we had put while creating the model. Matplotlib is used to plot the accuracy graphs, loss graphs and printing out the evaluation accuracy and evaluation loss. [9]

Implementing Speech Recognition

The above procedures were done for converting Sign language to Speech. Now coming back to Speech to Sign language conversion, we first need to implement the speech recognition to allow speech to be recognized. We will be using the speech recognition module for that. Once we receive the speech through microphone, it will be converted to text. After converting to text we will call the Gesture Display component or the Sign Display component after performing some validation checks.

Gesture Display and Sign Display

After receiving the text using Speech Recognition module, we will call the Gesture Display component if the text matches the gestures that we have in our dataset. If the gesture matches, then a gif image will be displayed to the user, indicating the gesture corresponding to the text which was spoken. If the text does not match with any of the gestures, then we will display a series of signs from the test data of the sign language dataset. The signs will be repeated twice having regular intervals so that user will be able to perceive them properly.

Implementing Live Classification

After creating the model for converting sign language to speech, we will use it to convert the signs provided through live web cam feed. We will be using OpenCV to get the images through the live feed. After receiving the images, we will pre-process them similarly as mentioned in section 5.2. After pre-processing, the image is given to the model to classify. The model will give us a numeric output corresponding to the alphabet that the sign portrays.

That alphabet is taken and printed over the screen for user accessibility. After the word is at least of 7 letters long, we will invoke the text-to-speech module to convert that text to speech.[10]

Providing A UI

The functionalities of sign to speech and speech to sign are decoupled here. To allow user to access them from a single point, we used the Tkinter library to have a GUI which had the options to convert from speech to sign, sign to speech and exit. Clicking an option would invoke the respective functionality of converting either sign to speech or speech to sign.

SYSTEM ARCHITECTURE

Sign to Speech

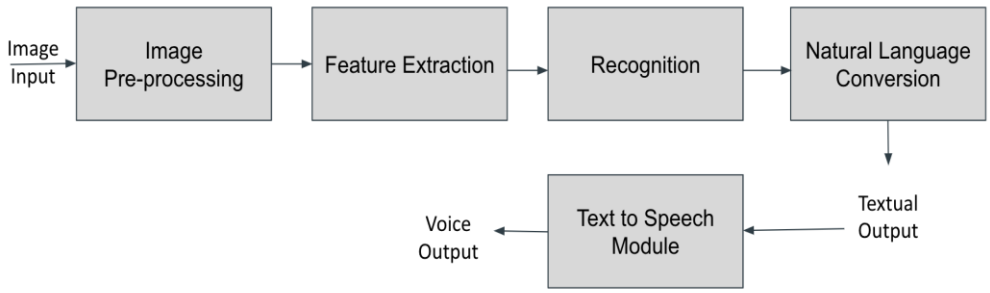


FIGURE 6: Architecture of Sign-To-Speech

Image Pre-Processing

Image processing can be defined as the processing of a digital image by with of a computer. It uses computer algorithms to get better image either to extract some useful information or for some other kind of task.

Feature Extraction

Feature extraction can be defined as reducing the dimensionality of an image in which the raw data is divided and reduced to more manageable groups.

Recognition

Machine-based visual tasks are performed by image recognition. Labelling is given to the images with meta-tags, to performing image content search.

Natural Language Conversion

In natural language conversion, the recognized image is converted into a English text and text output is given as input to the text-to-speech module.

Text-to-Speech

This module take text as input generated by the natural language conversion and gives the voice text as output.

Speech-To-Sign

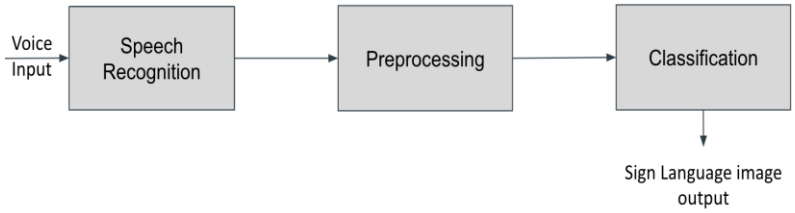


FIGURE 7 Architecture of Speech Sign-To- Sign

Speech recognition:

Speech recognition is a process of identifying speech by the computer. computers recognize the speech and convert that speech to the text. It is also called automatic speech recognition (ASR). Here the system will identify the speech spoken by the people and convert it to a text. That text is given as input for pre-processing.[11]

Pre-Processing:

Here the text undergoes pre-processing, the pre-processing involves identifying and converting the text to more system understandable format.

Classification

Here the classification means identifying the images of the signs based on the matched text. CNN will classify the images. When the speech is given the AI model will undergo classification and gives the image of the searching image.

RESULTS

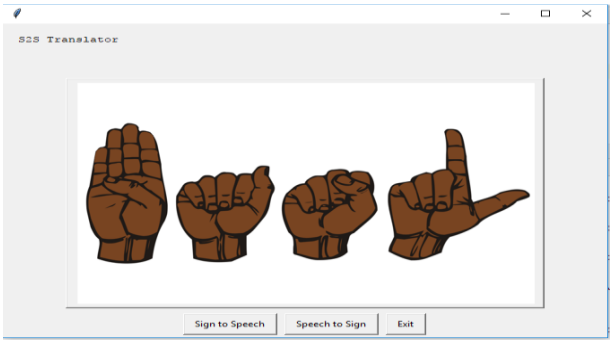


FIGURE 8.a Interface for user to choose options

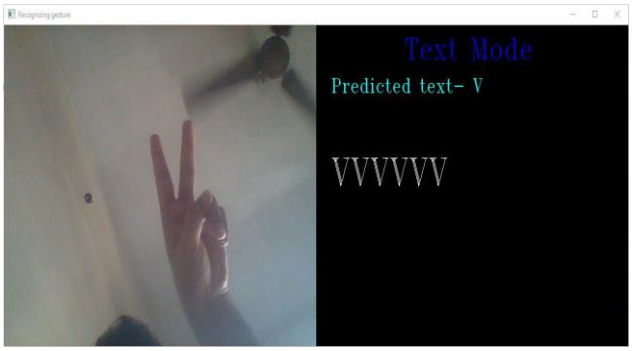


FIGURE: 8.b when user showing symbol 'V' and printing 'V'

FIGURE 8.a Here it is the login page, from this page we have to select the option convert to speech to sign or sign to speech.

FIGURE: 8.b For the sign we V the system showing it as Text form 'V' and also, we get voice also for the letter V

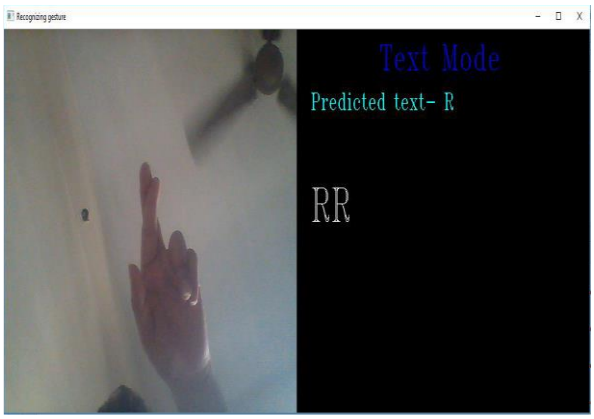


FIGURE:8.c for symbol 'R'

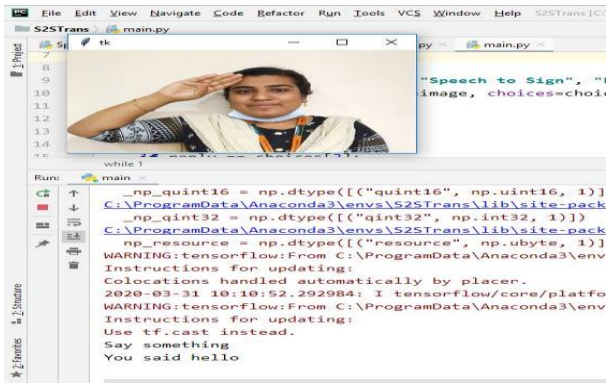


FIGURE: 8.d when user said 'hello' and system showing figure

FIGURE:8.c for the sign we R the system showing it as Text form 'R' and also we get voice also for the letter R.

FIGURE: 8.d This is the example for speech to sign, when the user said "hello", from the pretrained data the system showing the image for "hello".

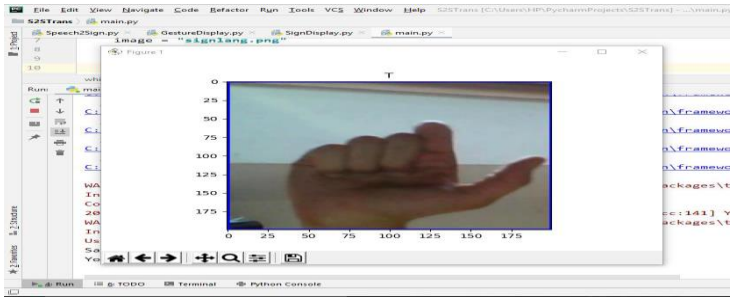


FIGURE 8.e: showing symbol 'T' when user said 'T'.

When user told the letter T from the pretrained data it showing as the letter "T"

CONCLUSION AND FUTURE WORK

The paper entitled "S2S Translator" has presenting a model for translating sign language to speech and speech to sign language. The model was developed for the fair enough accuracy and can be used for further improvisation. For better live classification, an object detection module could have been developed or an already existing one could have been used which will lead to better results in translating.

Some kind of support for translating more complex sentences can be further investigated for the future of this project. In Future should develop as application for this and should use through smart phone also. The voice output feature which is the output for the sign to speech conversion can be further improved.

REFERENCES

1. Anbarasi Rajamohan, Hemavathy R., Dhanalakshmi M."Deaf Mute Communication Interpreter", International Journal of Scientific Engineering and Technology (ISSN : 2277-1581) Volume 2 Issue 5, pp : 336-341.
2. V.Padmanabhan, M.Sornalatha," Hand gesture recognition and voice conversion system for dumb people", International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May-2014 427 ISSN 2229-5518
3. Jungong Han, "Enhanced Computer Vision with Microsoft Kinect Sensor": AReview, IEEE TRANSACTIONS ON CYBERNETICS.

4. Microsoft Kinect SDK, http://www.microsoft.com/en- us/kinectforwindows/.
5. Gunasekaran. K, Manikandan. R, International Journal of Engineering andTechnology (IJET): "Sign Language to Speech Using PIC Microcontroller".
6. RiniAkmeliawatil, Melanie PO-LeenOoi et al, "Real-Time Malaysian Sign Language Translation using Color Segmentation and Neural Network. Instrumentation and Measurement Technology Conference Warsaw",Poland.IEEE.1-6,2007.
7. Yang quan,"Chinese Sign Language Recognition Based On Video Sequence Appearance Modeling",IEEE. 1537-1542,2010.
8. Wen Gao and Gaolin Fanga, "A Chinese sign language recognition system" Journal of Pattern Recognition".2389-2402,2004.
9. Nicholas Born. "Senior Project Sign Language Glove", California Polytechnic State University,1-49,2010.
10. Kirsten Ellis and Jan Carlo Barca."Exploring Sensor Gloves for TeachingChildren Sign Language. Advances in Human-Computer Interaction".1-8.,2012.TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 59, NO.10,2695-2704, 2012.
11. Dey N.S., Mohanty R., Chugh K.L. "Speech and speaker recognition system using artificial neural networks and hidden Markov model" International Conference on Communication Systems and Network Technologies, CSNT 2012.
12. Koppula, N., Rani, B.P., Srinivas Rao, K., "Graph-based word sense disambiguation in Telugu language", (2019) International Journal of Knowledge-Based and Intelligent Engineering Systems, 23 (1), pp. 55-60.
13. Madhuravani, B., Murthy, D., "A novel secure authentication approach for wireless communication using chaotic maps", Proceedings - International Conference on Trends in Electronics and Informatics, ICEI 2017, 2018, 2018-January, pp. 360–363
14. Madhuravani, B., Chandra Sekhar Reddy, N., Sai Prasad, K., Dhanalaxmi, B., Uma Maheswari, V., "Strong and secure mechanism for data storage in cloud environment", *International Journal of Advanced Trends in Computer Science and Engineering*, 2019, 8(1.3 S1), pp. 29–33, 6