A comparative study on Sign Language Recognition Methods.

V. Venkataramanan[1]
[1]Assistant Professor
Department of Electronics and Telecommunication
Dwarkadas J. Sanghvi College of Engineering
venkatdjsce@gmail.com

Sarthak Dalal[2], Rohaan Kacheria[3]
[2,3]U.G. Student
Department of Electronics and Telecommunication
Dwarkadas J. Sanghvi College of Engineering
sarthak.dalal@gmail.com [2], rohankacheriam@gmail.com[3]

***Abstract* - The process by which two or more people exchange their information, views, and opinions on different situations and on different topics is known as communication. The need for a common platform for humans to communicate with one another has played an important role in the development of human civilization ever since its start. Human beings communicate with each other using several verbal and non-verbal cues. Non-verbal communication involves the use of signs and hand gestures. People who are hard of hearing or have speech impairments can communicate only using non-verbal cues and they rely on others to communicate with them using hand gestures as well. Systematic sign language systems such as the American Sign Language (ASL), the Indian Sign Language (ISL) and several such similar sign language systems have been developed to aid the specially-abled individuals. Sign Language recognition is a process that uses technology to recognize what the specially-abled person is saying by identifying the signs and hand gestures and returning the output in the form of text or speech. In this survey, a comparative analysis is done based on the various techniques and algorithms used and implemented to carry out the process of Sign language Recognition (SLR).**

Keywords: *Sign Language, OpenCV, Python, Machine Learning, Convolutional Neural network*

I. INTRODUCTION

Conversing with people who face hearing and speech disabilities has always been a tough task for people who do not face such disabilities. People with hearing disabilities and speech impairment use gestures and actions to convey information and their emotions to others. Parents are often the source of a child's early acquisition of language, but for children who are deaf, additional people may be models for language acquisition. A deaf child born to parents who are deaf and who already use a sign language will begin to acquire the sign language as naturally as a hearing child picks up spoken language from hearing parents. However, for a deaf child with hearing parents who have no prior experience with sign languages, language may be acquired differently. In fact, 9 out of 10 children who are born deaf are born to parents who hear. Some hearing parents choose to introduce sign language to their deaf children. Hearing parents who choose to have their child learn sign language often learn it along with their child. Children who are deaf and have hearing parents often learn sign language through deaf peers and become fluent. An interpreter is required to decode this sign language and to explain to others what the person is saying in a more lucid manner. The interpreter can also use sign language to explain to the specially abled person about the other person's responses. The interpreter therefore always has to be in presence of the specially abled and act as a bridge between others and the specially-abled. However, in the absence of an interpreter, conversing with people with speech and hearing impairment is not easy for others as the knowledge of sign language is not commonly known to the general public. Machine Learning and Deep Learning techniques have enabled the creation of technologies that convert Sign Language to Text or Voice format and vice versa. This enables the specially abled to converse with anyone without the need for an interpreter at all times. In this paper, we look at majorly two different categories of sign language recognition: Glove based or sensor-based approach and Vision based approach. Glove based approach requires the user to wear a glove embedded with various sensors that can extract the features of the hand and recognize what signs and gestures the user is making. This approach requires additional hardware which must always be with the user. On the other hand, in Vision based approaches, image processing, machine learning and deep learning techniques are leveraged to recognize the signs and gestures made by the specially abled in front of the camera. In this paper, we review the approaches for sign language recognition under these two categories and compare the advantages and disadvantages for both the methods. There are various different sign languages used all over the world, like the American Sign Language (ASL), Indian Sign Language (ISL), British Sign Language(BSL), French Sign Language (FSL), Chinese Sign Language (CSL), Arabic Sign Language, Turkish Sign Language, Brazilian Sign Language, Mexican Sign Language and more. All these sign languages have pre-defined gestures using either one or two hands that correspond to an alphabet or a word. In this paper, we will focus mainly on ASL and ISL systems and some of the technologies that have been used to develop the sign language recognition models.

II. DIFFERENT SIGN LANGUAGE SYSTEMS

A. *American Sign Language*

ASL is a complete, natural language that has the same linguistic properties as spoken languages, with grammar that differs from English. ASL is expressed by movements of the hands and face. It is the primary language of many North Americans who are deaf and hard of hearing and is used by many hearing people as well.[1]

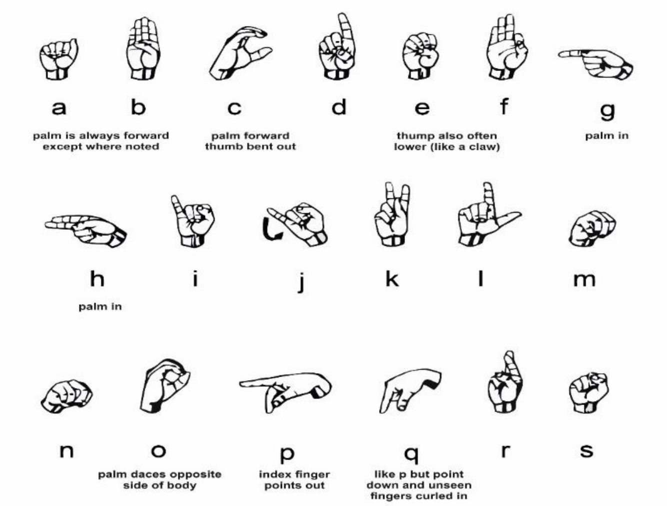

Fig. 1. American Sign Language gestures (Letters a-s)

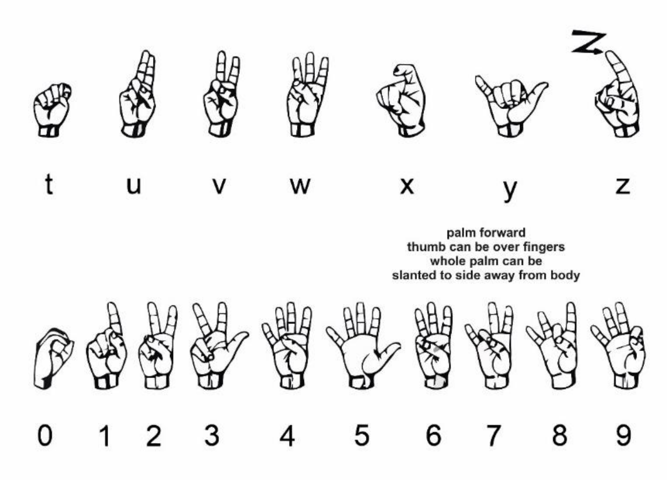

Fig. 2. American Sign Language gestures (Letters t-z and Numbers 0-9)

B. *Indian Sign Language*

Figure 3 shows a representation of hand signs in ISL. It can be observed that while some alphabets can be represented using a single hand, some alphabets require the use of both hands. This is a point of difference with the ASL as all the alphabets in the ASL can be represented with just one hand. Alphabets like A, B, D, E, F, G, H, K, M, N, P, Q, R, S, T, X, Y, Z use both the hands for their gestures representation. Alphabets like C, I, J, L, O, U, V, W make use of a single hand. Words and sentences can be conveyed using orientation of hands as well as facial expressions. More often than not, double handed gestures are used to convey words and phrases [2].

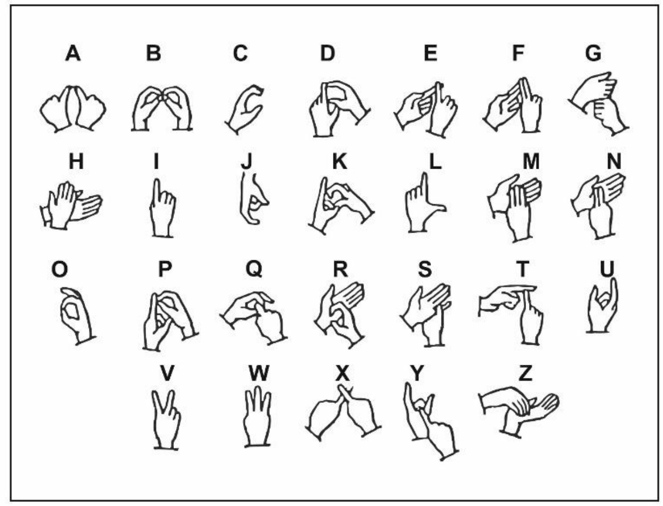

Fig. 3. Indian Sign Language gestures

[3] is a website dedicated to Indian Sign Language and it contains videos in which each alphabet, letters and many words of the English dictionary are shown using ISL.

III. MOTIVATION

The motivation behind developing this system is twofold, to learn and to communicate with others. Due to the lack of availability of systematic tools for the disabled people to communicate with one another and too pass on their messages to others, this system is developed so that the disabled people can lucidly communicate their messages to others without the use of any expensive external convertors or without being dependant on people who communicate their message to others. This system can be used not only for the disabled people to spread their messages but can also be used to oust language barriers by converting one language to another for example an Indian using the ISL can lucidly communicate to an American following the ASL.

IV RELATED WORK

Comparison papers which review different types of sign language recognition methods have been published. In [2], the paper also gives an explanation about the 2 different categories of sign language translators mainly being vision based and sensor-based systems. The first paper reviewed by the research is a sensor-based system in which a glove has to be worn by the user which is further used for capturing the gestures. The glove consists of the proximity sensor, accelerometer sensor, abduction sensor and many other sensors. The main use of these sensors is to extract features which describe the hand gesture. An android app could also be used to detect the hand gesture and the technology used to develop this system is based on wireless sensor technology as presented in the next category reviewed by them which is the vision based technique. A Kinect is used for capturing the gestures. Machine learning algorithms and an inbuilt camera is made use of to recognize the gesture. Digital Image processing technique is used to extract features of the sign and to recognize it.

[4] is another review paper which classifies the approaches in not two but three approaches – Glove based, Vision based and EMG (electromyography). The electromyography (EMG) sensor detects and records the electrical potential signal generated by muscle cells utilising differential pairs of surface electrodes. The paper later also classifies the methods using machine learning approaches such as Artificial neural network and Hidden Markov Model. Problems with the continuous sign recognition methods and future scope is discussed.

V. LITERATURE SURVEY

The problem of sign language recognition has been addressed using different methods by researchers. There can be two approaches to SLR - Glove or sensor-based approach and vision-based approach. The sensor-based approach can provide good accuracies as the sensors collect data and features directly from the hand of the specially abled. The downside is that the user must wear a glove with sensors attached to it all the time to communicate. If the user does not have the glove or the hardware malfunctions, then the user cannot communicate using this technique. Vision-based approaches use computer vision and various image processing techniques to recognize what the specially abled is trying to communicate. This approach can be further classified to 2-D and 3-D vision-based techniques.

A. Sensor Based Approach

In a sensor-based approach, the setup usually consists of a receptor or input unit, a processor, and an output unit. Sensors such as flex sensors, accelerometers, IMUs etc are used as the input unit to measure hand orientation, movement, bending of the fingers etc. The processor is a microcontroller which processes the data from the sensor for recognition or transmits it to a higher-level computer for the recognition process. The gestures can be recognized using several different machine learning techniques.

1)Artificial Neural Network

An Artificial Neural Network is a computational system based on the neural networks that are the basic units of animal's brains. Like the human brain which has neurons interconnected to one another, Artificial Neural Networks have similar neurons that are interconnected to one another in various layers of the networks. These neurons are known as nodes.

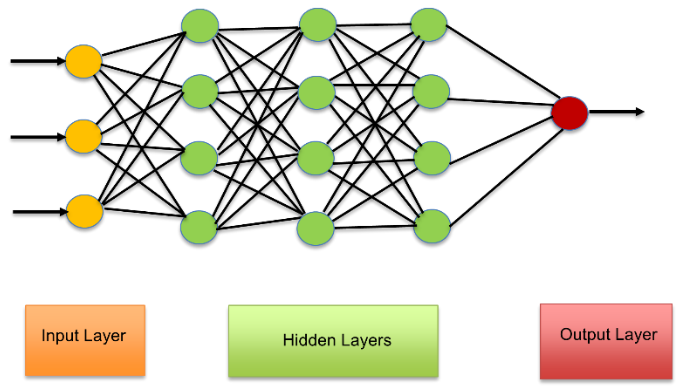

Fig. 4. Artificial Neural Network

An artificial neural network can be used to train and test a model for sign language recognition and the data can be fed to it by using a sensor-based approach. The authors of [5] use a glove for this purpose. The user must wear a glove which records the gestures and signs created by the user. The glove comprises of four flex sensors which detect the gestures. The resistance of the un-flexed sensor is 10k and on bending the fingers, the resistance of the sensors changes accordingly. The change in voltage accordingly is measured. This value is sent to the processing unit which is the Arduino UNO microcontroller. The flex sensor values are displayed on the Serial monitor of the Arduino IDE. A SIM900A GSM module is interfaced with the microcontroller and the data is sent onto a publicly available thingspeak cloud server. This is where the value of the data is plotted using a model trained on MATLAB by the MATLAB Neural Network Toolbox. This toolbox can implement various algorithms and methods such as back-propagation Neural Network, Particle Swarm Optimization and many more. The SIM900A module also sends an SMS to the receiver's mobile number. An android application is developed for this purpose which converts text to speech. When the user gets a push notification, the final output is displayed in the android application and the inputted text is converted to speech and the output is returned through the speaker of the mobile device.

2) Non-ML Approach

In [6], the performed gesture is recognised using 8 independent capacitive touch sensors which return an ON or OFF binary output. Five of these sensors are placed at the fingertips and three additional sensors are placed between the index, middle finger, ring and pinky finger. These sensors get triggered when they are within 1.6mm radius of human skin. Since the sensors could be triggered by materials with a dielectric constant lower than that of human skin, a polymer type of material is used to develop the glove. The American Sign Language letters A-Z and the number 0-9 are mapped to specific combinations of the 8 sensors. The advantage of this is that the signal sent is unambiguous and it is easier to predict which sensor is active during a particular gesture, hence reducing any mistakes for prediction. All of these 8 sensors are implemented using the PIC-116 capacitive sensor module having a QT100 capacitive sensor integrated at its core. Binary data from the touch sensor is transmitted to integrated Analog to Digital converter (ADCs) of the Raspberry Pi. The main function of the firmware is to recognize the gestures which are performed using the lookup process. In order to allow the user to complete the entire gesture, a 3 second countdown is implemented before reading each gesture. The prototype has short start-up and response times and can return a result in 0.7 seconds with most of the processing time being spent on loading the audio file. There were a total of 36 gestures, 30 trials were done on all gestures leading to a total of 1080 trials. This computation led the authors to get an accuracy of 92%.

The main aim of [7] is to develop a sign to a letter translator system. The glove is made accessible to a deaf and/or mute individual to communicate with people who do not understand sign language. The signs identified are displayed as letters on an LCD screen. The system consists of six flex sensors held by the glove, discrete components, microcontrollers and LCD screen. The microcontrollers are the main component of the system and it analyses the changes and returns the output on the LCD screen. The input is the sign from the flex sensor which is connected to a power supply and a circuit. The microcontroller is operated alongside with a 4MHZ clock to provide enough data to its connected components as well as to receive accurate data. It also contains a 10-bit 8 channel analog to digital converter. Some of the additional components in the PCB are the oscillator and a voltage regulator which converts the 9 volts supplied by the used battery to 5 volts which is needed for the optimal functioning of the microcontroller. The program starts by reading the analog inputs of the flex sensor and reserves 2-bit positions for each flex value. If else conditions are made use of alongside functions to return the desired output. The voltage values of each of the 26 alphabets were taken in advance since the resistance value of each alphabet was different. The system developed has achieved an average accuracy of 94% when trying to display all letters.

In [8], the system implemented consists of 5 flex sensors and an accelerometer which is made use of to identify the hand and palm movements at the input. The accelerometer detects the rotation of the palm which helps choose the language in which communication is to be done. The resistive strips of the flex sensors are folded to vary the resistance and to return either a 0 or 1 logical output to the inbuilt ADC of ARM7. This way a 6-bit binary digit is formed for each gesture that is performed. Due to the use of 5 similar flex sensors, 32 hand gestures are possible to be implemented which is stored in the external memory. The microcontroller than controls this reading to the lookup table and the digit with the closest value is then returned. The ARM7 then searches the external memory for audio files commensuration of the binary digit. If the user gesture is showing the same gesture, ARM7 will go on to find the binary digit of the word and then the audio file which is named after the word. If the file exists, the file will be played with the correct pronunciation.

B. Vision Based Approach

Vision based approaches are more favoured over sensor-based approaches in sign language recognition. This is because of the fact that using a vision-based approach eliminates the need to depend on sensors, processors and extra hardware, thereby also reducing the overall cost. It is also more reliable as the approach is software based rather than hardware based.

1) Convolutional Neural Network

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an image, assign weights and biases to various aspects of the image and is able to differentiate one bias from the other. These algorithms can identify faces, street signs, individuals, tumors, and other aspects of visual data. Convolutional networks can perform Optical Character Recognition (OCR) to digitize text and make Natural-Language Processing possible on analog as well as hand-written documents, where the images are symbols to be transcribed.

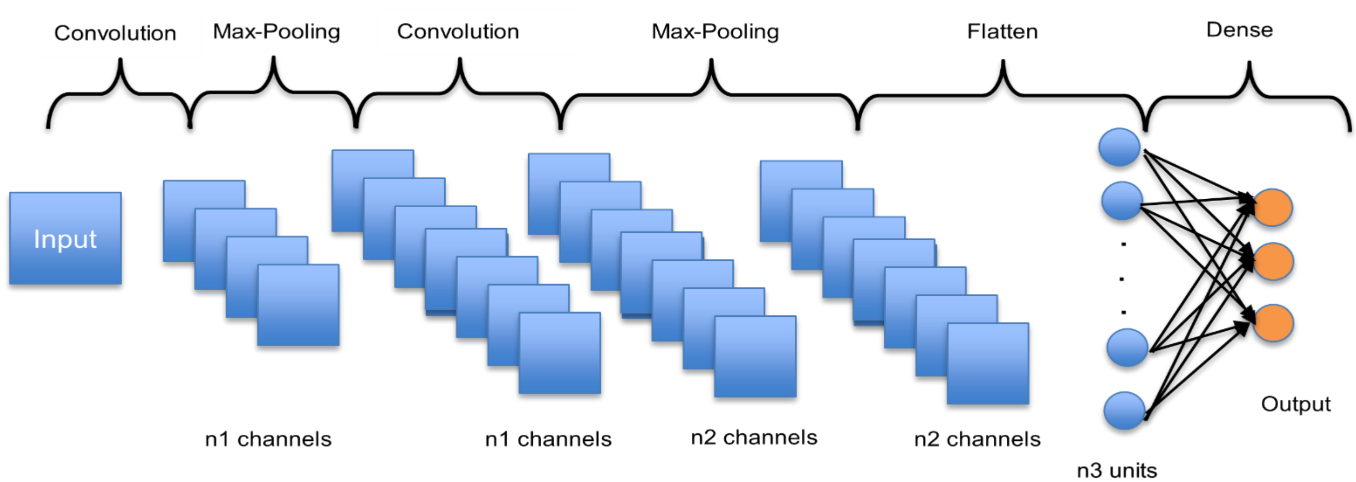

Fig 5. Convolutional Neural Network

A CNN model is used in [9]. A dataset of 1000 images for each of the 33 different hand signs (letters and numbers) of ASL from 5 different individuals has been collected using a Creative Senz3D depth camera. The resolution of this camera is 320x240. Letters J and Z are eliminated as they require temporal information. 2/V and 6/W use similar gestures and are differentiated based on context; hence a single class is used to represent them. The Python Deep Learning Toolkit under the Theano environment was made use of. A single CNN with 20 feature maps, local filters having a size of 5x5 and a pooling size of 2x2 was made use of. The number of target classes returned were 33. The FC hidden layers ranging from 1 to 4 were made use of alongside a learning rate of 0.1. 784 features were extracted from each of the pre-processed depth images. The images were then segregated from training, validation and testing purposes. The Neural Network was trained by varying the hidden layers from 1 to 4 and by varying the epochs from 500 to 1000. The number of nodes varied as well. Out of the 1000 images, 900 were used for developing the training set, 60 for validation and 40 for the testing set. A simple JAVA GUI was developed to test the classifier on static sign language. The application allowed users to choose words or letters and they were then spoken out using the python pyttsx text to speech module. Multiple combinations of the number of hidden layers and epochs were tried to maximise accuracy and the best accuracy was 94.667%.

The authors of [10] decide to use more images for training the CNN. The data is based on images from ASL only, which uses a single hand, precisely the left hand. The positioning of the hand is also not in the centre and the hand covers all possible angles. This helps in increasing the accuracy. The data consists of 87,002 images and is split in the ratio of 9:1 as train and test data. The data was preprocessed by converting the pictures to grayscale images as grayscale images contain just one shade, which is easier for the CNN to learn. Normalization was done by dividing with 255 so that the images now lie between 0 and 1. The images are resized to 100*100. After the images are normalized and resized, they are fed to the CNN for training. The model has two convolutional layers and two max pooling layers. The ADAM optimizer function is used. The activation function used is the Rectified Linear Unit (ReLU). The chosen evaluation metric to train the model on is Accuracy. The output layer has 29 nodes which corresponds to the 29 classes of the output. The training accuracy achieved is 98.68% and the validation accuracy is 90%.

[11] In some scenarios the dataset is not available readily and the dataset has to be manually prepared. The authors of the research paper prepared the used dataset manually for the Indian Sign Language for their Convolutional Neural Network model to train on. Every class of images has a minimum of 2000 images in the dataset. The Dataset is created using different backgrounds to return a higher accuracy. The dataset also accounts for variations in the hand gestures and hence, several images are inserted with movement in the hand gestures.

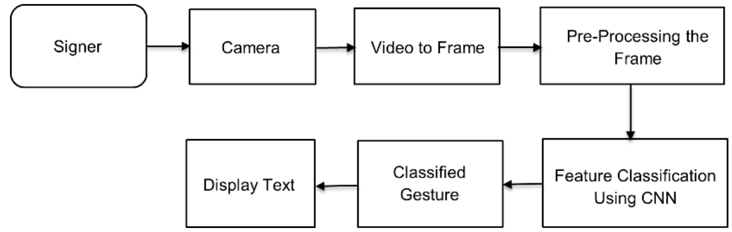

Fig. 6. Flowchart of the Methodology

Image processing was done using OpenCV which allowed conversion of video to frames further allowing the downscaling of images and converting the images to grayscale so that the skin tone has no effect on the final output. The You Only Look Once (YOLO) object detection algorithm was used to aid with hand detection. A CNN was used for gesture classification. The VGNet6 CNN algorithm was made use of. It gave an error rate of 7.3%. A few defects were identified and stated in the paper hoping to oust them in the future. One of the defects faced by the model was the variation in the skin tone of different individuals. Whilst working with different skin tones, the model would drop the accuracy. Similar problems would arise when the input image would contain the face of various individuals leading to a further fall in the accuracy of the model. This is why they trimmed the videos to only include hand gestures and no other features.

Another feature that can be added is the "text to speech conversion", so that the feel of a real conversation can be implemented. The authors of [12] do that by implementing the python "pyttsx3" library. A webcam is used to capture images of a man or a woman's hand signs. The image is pre-processed and fed to a keras CNN model which gives the prediction in text. Figure 7 explains the flow of the methodology.

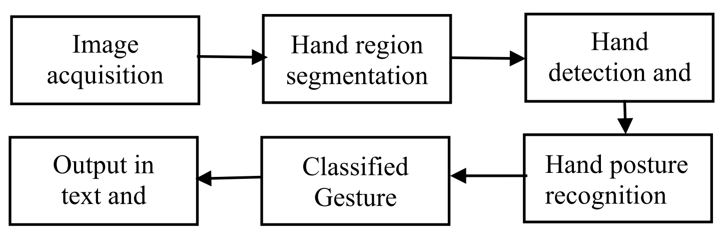

Fig. 7. Flowchart of the Methodology [12]

The CNN model consists of 11 layers. There are 3 convolutional layers and the first layer accepts images in grayscale of dimension 50x50. 16 filters of the size 2*2 are used in this layer which results in the generation of an activation map of the size 49*49 for each filter which means that the output is equivalent to a dimension of 49*49*16. Negative values are removed on the map and replaced by 0 using a Rectifier Linear Unit (ReLU) layer. Three max pooling and a dropout layer is present in the model architecture. A dropout layer is used to drop out random map

elements to minimize overfitting. A dense layer, in the end reduces the map to an array of 44 elements. This represents the number of classes. Each class has a corresponding probability of prediction that has been allocated to it. The class with the maximum probability is displayed as the final predicted gesture. There are six steps in the algorithm of the project. The first step is adjusting the hand histogram according to the lighting conditions and skin complexion. Then data augmentation is carried out on the dataset to expand it and thereby, minimize the overfitting. The dataset is then split to train, test and validation sets. In the fourth step, the CNN model is fit onto the dataset. Next, the model reported is generated alongside the confusion matrix which gives the accuracy and error. Finally the prediction file is executed which gives the output to hand signs in the form of audio and text.

2) Linear Discriminant Analysis

Linear discriminant analysis (LDA), normal discriminant analysis (NDA), or discriminant function analysis is a generalisation of Fisher's linear discriminant. Fisher's linear discriminant is a method used in statistics and other fields to find a linear combination of features that characterises or separates two or more classes of objects or events. The resulting mixture can be used as a linear classifier or, more typically, to reduce dimensionality before further classification. The most common applications of LDA include statistics, pattern recognition, and machine learning.
The author of [13] uses the LDA approach. The paper deals with the 26 hand gestures used to represent alphabets in ISL. In the data, 10 images for each 26 signs are taken and are included in the training and testing database. Segmentation is the process of dividing a picture into small segments to extract the more accurate image attribute. Segmentation on the images of the hands are carried out to separate it from the background. Next, dilation and erosion functions are performed on the images. Dilation increases the number of pixels on the edges of objects in an image, whereas erosion reduces the number of pixels on the edges of objects. Principal component is used as the main feature for feature extraction. For the sign recognition, dimensionality reduction takes place which is a technique based on extracting certain number of principal components from the multi-dimensional data. Figure 8 gives the flowchart for the methodology used

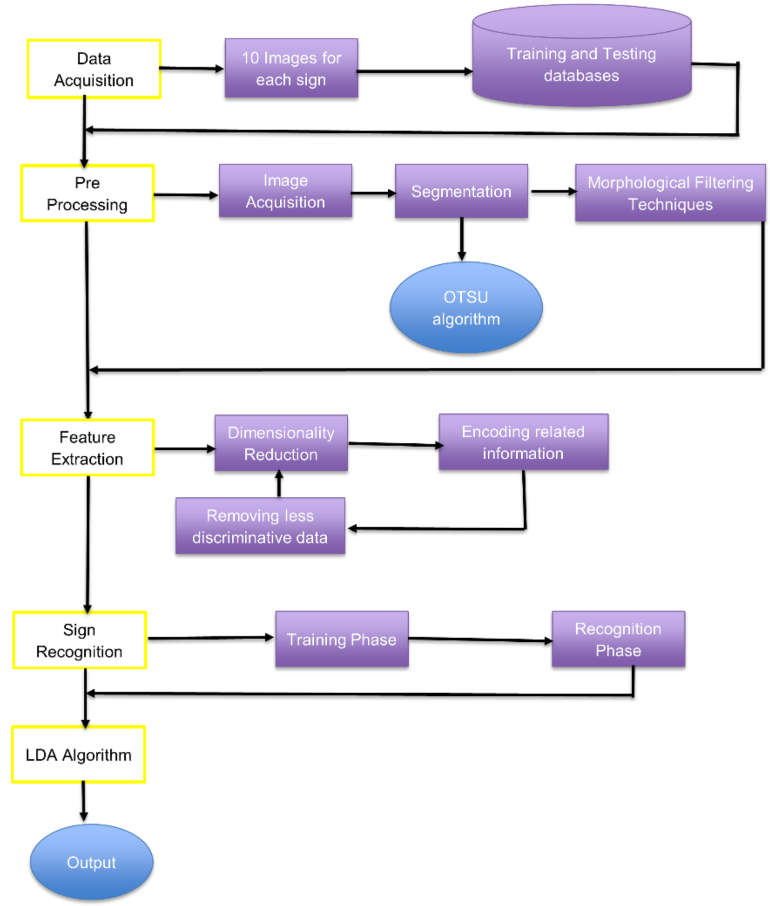

Fig. 8. - Overview of the methodology used [13]

A subject gesture is normalised with respect to the average gesture and then projected onto gesture space using the eigenvector matrix during the recognition phase. Finally, the Euclidean distance between this projection and all other projections is calculated. During the training phase, the minimum value of these comparisons is chosen for recognition. Finally, the identified sign is translated into appropriate text and speech for display on the GUI.

3) Artificial Neural Network

Artificial Neural Network models can also be used for vision-based recognition. In [14], The database used for the research was self-made due to the lack of Indian sign language numerals on the internet. The database comprises 1000 images, 100 for each numeral. Data reprocessing was then carried out which helped the researchers oust the redundant and unwanted data leaving behind the more informative data. To recognize skin shading, the image is changed to YCbCr shading space. This produces a binary image and is one of the most popular methods for extracting the skin portion of an image. In order to categorize the gestures from the binary image, they needed to take out the definite characteristics of the image. This was done by carrying out multiple feature extraction techniques, namely

HOG, shape descriptors, SIFT, PCA and Fourier descriptors. These features were then given as the input to the next classification step. The SVM model and ANN were used for classification purposes. The SVM model coupled with various feature extraction models gave a 93% accuracy whilst the ANN model coupled with the HOG gave an accuracy of 99%. This entire process was carried out on MATLAB.

In [15], the authors use the leap motion controller for input. The Leap Motion controller is a small USB peripheral device that allows users to operate their computers only through hand movements. This sensor is a three-dimensional non-contact motion sensor that detects and tracks hands, fingers, bones, and finger-like objects, reporting discrete position and motion. Two monochrome IR cameras and three infrared LEDs are at the heart of the device. The authors work on only recognising the gestures for numerals (0-9). A dataset of a total 200 images is created (20 images for each number). The training set is made of 100 images and the rest 100 images are used for testing. Before sending the data to the ANN for training and testing, it is normalised. Training then takes place on a multilayer perceptron neural network. The MLP consists of three types of layers input, output and hidden layer with one or multilayer. The result was outstanding as 100% accuracy was obtained on the remaining 100 images (10 images for each number) in the test set.

VI CHALLENGES AND FUTURE SCOPE

Sign language recognition techniques do not have to be restricted to just ASL and ISL systems. Such techniques can be developed for sign language systems all over the world. The specially abled use their expressions, elbows and even their whole body to convey emotions, hence only hand sign language recognition will not be sufficient. Recognition of facial expressions, elbow movement and complete body movement can be a scope for the future. In a video conference call, live translation of sign language to voice and text and vice versa can be implemented so that the specially abled can lucidly participate in such calls.

VII. CONCLUSIONS

In the path of establishing a robust communication system between the impaired hearing people, several approaches and gestures recognition systems have been developed. Apart from the American Sign Language (ASL) and the Indian Sign Language (ISL), many other such systems have been developed for efficient communication. Many such systems are undergoing progress with various classification algorithms. In this paper we looked at mainly two classes of sign language recognition – the sensor-based approach and the vision-based approach. The sensor-based approach can be tackled without the need for an artificial intelligence or machine learning model, which makes it a convenient approach to choose. The limitation however is that this approach requires the use of additional hardware such as sensors and microcontrollers, attached to the wearable device of the user. If this hardware malfunctions or is not with the user at the time of need, this system entirely collapses. On the other hand, in the vision-based approach, entire system can be trained on software only and a single camera is required for recognition of the sign language. Algorithms and machine learning models can be trained for sign language recognition. The advantage is that there is no reliance on special hardware and the software can be used on multiple platforms. Hence this system is more robust compared to the sensor-based approach. The future scope and the potential challenges that we could face from these systems are constantly being detected and being worked on to get rid of. More work needs to be done on developing such systems so there is just one system functioning harmoniously and can be easily available to everyone.

REFERENCES

[1] "American sign language," National Institute of Deafness and Other Communication Disorders. [Online]. Available: https://www.nidcd.nih.gov/health/american-sign-language.

[2] Sharma, Swati, Varun Prakash Saxena, and Kandukuru Satish. "Comparative analysis on sign language recognition system." Int. J. Sci. Technol. Res. 8.8 (2019): 981-990.

[3] "Welcome to indian sign language portal," ISL Indian Sign Language Portal FDMSE RKMVERI COIMBATORE CAMPUS. [Online]. Available: https://indiansignlanguage.org/.

[4] Al-Ahdal, M. Ebrahim, and Md Tahir Nooritawati. "Review in sign language recognition systems." 2012 IEEE Symposium on Computers & Informatics (ISCI). IEEE, 2012.

[5] A. Abraham and V. Rohini, "Real time conversion of sign language to speech and prediction of gestures using Artificial Neural Network," Procedia Computer Science, vol. 143, pp. 587–594, 2018.

[6] K. S. Abhishek, L. C. F. Qubeley and D. Ho, "Glove-based hand gesture recognition sign language translator using capacitive touch sensor," 2016 IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC), 2016, pp. 334-337, doi: 10.1109/EDSSC.2016.7785276.

[7] H. El Hayek, J. Nacouzi, A. Kassem, M. Hamad and S. El-Murr, "Sign to letter translator system using a hand glove," The Third International Conference on e-Technologies and Networks for Development (ICeND2014), 2014, pp. 146-150, doi: 10.1109/ICeND.2014.6991369.

[8] Priyanka Lokhande, Riya Prajapati and Sandeep Pansare. Article: Data Gloves for Sign Language Recognition System. IJCA Proceedings on National Conference on Emerging Trends in Advanced Communication Technologies NCETACT 2015(1):11-14, June 2015. Full text available.

[9] Beena, M. V., MN Agnisarman Namboodiri, and P. G. Dean. "Automatic sign language finger spelling using convolution neural network: analysis" Int J Pure Appl Math 117.20 (2017): 9-15.

[10] P. V. S. M. S. Kartik, K. B. V. N. S. Sumanth, V. N. V. S. Ram, and P. Prakash, "Sign Language to Text Conversion Using Deep Learning," Inventive Communication and Computational Technologies, pp. 219–227, Sep. 2020.
[11] Patil, Uday, et al, "Sign Language Translator Using Deep Learning," No. 2541. EasyChair, 2020.
[12] Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur, Dr. Dayananda P, 2020, Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCAIT – 2020 (Volume 8 – Issue 15).
[13] NB and Mahesh Kumar "Conversion of sign language into text." International Journal of Applied Engineering Research 13.9 (2018): 7154-7161.
[14] J. Ekbote and M. Joshi, "Indian sign language recognition using ANN and SVM classifiers," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Mar. 2017.
[15] D. Naglot and M. Kulkarni, "ANN based Indian Sign Language numerals recognition using the leap motion controller," 2016 International Conference on Inventive Computation Technologies (ICICT), 2016.

VIII. ABOUT THE AUTHORS

Sarthak Dalal: Sarthak is a final year student, pursuing Bachelor of Engineering (B.E) in Electronics and Telecommunications Engineering from Dwarkadas J Sanghvi College of Engineering, Mumbai. He is a process oriented individual with a flair for programming and his interests lie in Artificial Intelligence and Machine Learning, Internet of Things and Robotics.

Rohan Kacheria: Rohan is a final year student, pursuing his Bachelor of Engineering (B.E) in Electronics and Telecommunications Engineering from Dwarkadas J Sanghvi College of Engineering, Mumbai. He has got a strong grit for programming and aims to use data to bring about meaningful changes to the society. His interests mainly in the field of Artificial Intelligence and Data Science.