International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP. 181-188

Conversion of sign language to text for deaf and dumb

Poorvi Desai[1], Preksha Mohre[2], Soumya Bani[3], Varada S Naik[4], Sudha P.R[5]
[1](Dept. of ISE, JSS Academy of Technical Education - Bengaluru, India)
[2](Dept. of ISE, JSS Academy of Technical Education - Bengaluru, India)
[3](Dept. of ISE, JSS Academy of Technical Education - Bengaluru, India)
[4](Dept. of ISE, JSS Academy of Technical Education - Bengaluru, India)
[5](Dept. of ISE, JSS Academy of Technical Education - Bengaluru, India)

Abstract: *Sign language is considered as one of the oldest and most natural forms of communication. But most people are not aware of it and translators are difficult to find, so came up with a synchronal approach for gesture recognition based on American Sign Language by utilizing neural networks. The new inventions and advanced Computer Vision based technology has made utilizing of sign language less complicated and much easier. The hand gesture is firstly filtered by using gray scale and Gaussian filter properties to distinctly identify the gesture. Then the process goes through the classifier for identification and filtering that predicts the motions made by hand. The result from the recognition of hand gesture will be converted to text and displayed onto the output. This project will be a great help to the hard of hearing to communicate with ordinary people.*
Keywords - *language, Computer vision, neural networks, Hand gesture, Gaussian filter, Grayscale*

I. Introduction

American sign-language is considered as one of the most common and appropriate sign language. As deaf and mute people cannot use spoken languages, sign language is the only option for their communication. The sign language is a form of communication where gestures represent some actions through which they convey their messages. These sign languages are only known to deaf and mute people and the translators. The ordinary people who are unaware of the sign language find it difficult to communicate with them. So to eradicate this problem, an effort has been made to come up with a project "Conversion Of Sign Language To Text".

II. Literature Survey

In the paper [1] have implemented Sign language for dumb and deaf community as rest of the ordinary people verbally communicate with each other but dumb and deaf due to their inability to talk or hear cannot verbally communicate with each other. This paper explains how they implemented a model or a system that assistance dumb and deaf communities. This study intends to close the breach in communication by presenting a unique method for interpreting stable and vigorous signals in Sign or hand gestures and converting it to voice form. This information is then wirelessly sent and categorized into voice outputs. Because of its capacity to learn long-term dependencies, these networks were explored and applied for hand recognizing data categorization. The proposed model could categorize 26 movements with a 98 percent accuracy, demonstrating the viability. Methodology: In the 2011, nearly 7 million Indians are having hearing-impaired and dumb, with 1.9 million having talk impairment. Sixty-three percent are unemployed, and thirty percent have never attended school. They can only converse with each other via signs and gestures due to their disabilities. These people communicate using hand gestures, the remainder of the population communicates vocally. This causes a schism in communication. As a result, they are at a downside and have no access the similar schooling and work prospects. This is a collection of hand gestures movements utilized by India's deaf people. These motions are divided into the following types: stable and vigorous. Stable gestures include the hands that do not move while making the signs. The majority of ISL's alphabet motions are stable indicators. Vigorous gestures, include entail finger motions while executing signs and account for the vast large number of hand signs. A technology competent of converting these gestures to voice might be effective, allowing the deaf to converse with ordinary or normal community. Their social lifestyle of communicating with each other will be changed. The gadget must identify both static and dynamic motions. While typical these can categories stable movements, which are not classify dynamic motions. Recurrent neural networks may also describe sequence and temporal dependencies on several scales at the same time. However, it has been found that when the time range of dependencies rises, rising ground subsiding of the mistake criteria can become a model which is not very flexible and inefficient

in its performance. These dependencies, can be used to overcome this inefficiency. This network design which can viaduct temporal spans more than thousand steps paradoxically when the input sequences are noisy and incompressible. Because of its capacity to study of these dependencies, the networks can be utilized to categories data of sensor matching to the gesture motions which is related the labels, assisting in the development of a device capable of real-time translation of sign language. They demonstrate the model and operation of a hand glove which can be recognized by camera and translates signals to voice in this study.

In the paper [2] proposed a system which is mainly based on deep learning to get an accurate result. For American Sign Language (ASL) gloss detection and reception, system use open cv, respectively. They showed in this research that an throughout the system has the potential of apart from being able to recognizing singular hand movements but also it is able to convert these words into phrases as it is more efficient to understand the process. Using fingers to demonstrate their words or phrases of any kind of communication these have distinct movements for a particular letter, is an alternative approach to recognizing sign languages. One frequent method is saving a hand gesture or any movement later when the system or model is run the process will compare it with the saved image and will provide the effective result [9]. The only difference in hand movement and contemplate Finger- spelling identification systems, external devices can be utilized for brief dialogues and basic gesture based inputs. It is, nevertheless, impracticable for everyday conversation. In practice, sign language incorporates diverse movement, which includes complicated words, every single letter, and movements that lack equivalent signals. Dept of ISE, JSSATEB 7 A. Stage 1 of this project includes the recognition from a video. In this video, the user will be displaying or making hand gestures and the system needs to recognize that particular hand gesture. To achieve this goal, the isolated gloss identification system employs by recognizing from the video of the user which is the continuous movement of the hand and includes the module with a network. Stage 2 - Neural Translator for the video to Speech The collection of these images captured from a video determined as in the previous step which translated into an untestable phrase in this stage. IV. DATA SOURCE – From the resources available it provides the films needed to test the data which is required for the translator system or model. For training and testing, only sensitive movies with precede sight gestures are used. V. In this stage the system recognizes the images from a video preparation. In this system, it recipes a continuous movement of the hand gestures or a video of end user performing some sort of hand movements or executing motions which is captured by the system camera as input. The noise in the background is not removed by this system so the video must be captured when the user must be in dark clothing with no or low noise in the background and a background which must be in dark color. The system analyzes the video by analyzing each image captured by the camera with each picture being treated as source to the system for further analyzing and processing of the image to provide an accurate outcome. In the initial stage, the hands of the user and other unwanted images captured by the camera are removed accordingly. This step is done as the hands of user gives majority information from the captured video and provides an effective result. This process is done through the use of the Gaussian Blur filter and Otsu's Binarization. The pixel in the captured image will be adjusted by Gaussian Blur Filter and it also smoothens the image(frame).

This paper [3] states that due to the intricacy of signals and limited data gathering, vision-based sign language recognition is a difficult undertaking. The research presents HMMs (Hidden Markov Models), adaptive GMM-based (Gaussian mixture model) framework to increase recognition precision. The paper also shows that the count of important gestures and body positions in HMMs is connected to the amount of underlying latent states and the translation linkage type. The paper presents use of adaptive HMMs and propagation clustering affinity to calculate each sign's hidden state number. Furthermore, it presents a data augmentation technique that includes Gaussian random disturbances to enrich the training dataset. Experiment are conducted on gestures of almost 370 signs to show the out performance of the technique being used in comparison to other methods. Implementation: SLR (Sign Language Recognition) has piqued the interest of researchers in pattern recognition and computer vision. Due to the intricacy of the symptoms and the restricted data collection, it remains a difficult work. Many studies were done on tiny datasets due to inadequate data collecting. One example is Kurakin et al's suggestion of a real-time SLR system which uses a dataset consisting of 12 American Sign Language (ASL) movements. Dong et al. got a recognition precision of 90%. Sun et al. recognised 73 ASL signals with 85.5 percent accuracy. Two cutting-edge works in huge vocabulary SLR are launched based on research. Ong et al. offered one that obtained 74.1 percent accuracy on 982 signs, while Wang et al. proposed another that reached 94 percent accuracy. The top ten had an accurate sign accuracy of 78 percent. The dataset was annotated to indicate whether the sign is single handed or double handed, as well as which hand is dominant. It made advantage of earlier sign knowledge. In actuality, it is impossible to predict who the signer is. Furthermore, in order to obtain a

powerful modelling capacity, researchers use several models such as Curve Matching (CM), Dynamic Time Warping (DTW), Hidden Markov Models (HMMs), and neural networks which are known to provide accurate and trustworthy results in dynamic speech recognition. Method of Dynamic Time Warping (DTW) with a linear space and time complexity are used on gesture recognition. Lin et al. suggested a curve fitting approach with sign trajectory based on manifold analysis. Wang et al. presented Light HMMs for selecting primary frames using lower rank approximation and determining the count of states that are hidden using a RSS(Residual Sum of Squares) threshold method. There are also an increasing number of recurrent neural networks (RNNs) utilized to handle the SLR problem. In general, previous knowledge about signs or signers is difficult to get in a real-world application setting. Furthermore, because sign language knowledge is less and data collection with depth is uncommon, there exists a scarcity of enough training examples. Improving the recognition precision of SLR remains a difficult task. Given the HMM's exceptional capacity, an approach for data augmentation is provided. The paper introduced random variables of noise with Gaussian distributions into the original data. The amount of states present in every sign of HMM model highlights crucial action and gesture Dept of ISE, JSSATEB 9 modifications in the SLR problem. The paper presents an HMM-state adaptation technique based on clustering of affinity propagation to identify the optimal state count adaptive. From our literature survey, it is inferred that deep learning techniques and openCV have greater impact on detection and conversion of sign language to text. In addition, CNN architectures are popular for the task of image processing. In sign language recognition, image processing is used to better extract features from input images. Images are in static image or dynamic image of sign performed by human. Gesture recognition is fast growing field in image processing and artificial technology.

In paper [4] there are very few sign learning study resources accessible. As a result, the process of learning sign language is extremely tough. Finger written sign learning is the first step of sign learning and is utilised when no equivalent sign exists or the signer is unaware of it. The majority of available technologies for sign language learning rely on expensive external sensors. Our study intends to advance in this sector by gathering a dataset and then employing various feature extraction algorithms. A. Photo capturing is the initial stage in learning to recognize signs. Camera interfacing is a vital component. The hand motion is captured using a web camera. This web-camera may be an internal camera embedded into computer or an external camera. The captured photographs must be of good quality and of high definition. Thus, selecting the right webcam and interfacing it is a crucial step used in this technique. B. Image Preparation Cropping, filtering, brightness and contrast adjustments, and other techniques are all part of image preprocessing. Image enhancement, cropping, and segmentation technologies are employed in this procedure. Captured images are in RGB format. The initial step is converting RGB photos to binary images. The image is then cropped to remove any unwanted part. C. Feature Extraction- It is a highly helpful stage in creating a database of sign language for the project. To economically and efficiently define the varied visual principles in the manual alphabetic letters, both global and local visual elements are retrieved for characterization of similarity between letter and image. In sign recognition, primarily two kinds of feature extraction approaches are employed: i) based on contour which comprises of description and presentation methods and ii) based on region which comprises of shape description and Dept of ISE, JSSATEB 13 presentation methods. The ones based on application techniques are chosen from among those available. The 7Hu moments strategy is applied in this suggested method, and 7 moments are identified as a result of that. From those times, agesture database was created.

In the paper [5] have implemented the Hand sign recognition .This has been a promising topic that has been used to a variety of practical applications: Surveillance cameras, for example, detect and recognise hand gestures in order to prevent illegal activity.A range of research have also looked at hand gesture recognition, including sign language recognition , lie detection and robotic control. The variables in an image area is so huge for a human hand gesture recognition system on the basis of images. It is critical to extract the picture's key properties. A large training database is often necessary to develop an effective hand gesture recognition system, and diverse motions should be represented. They have built a human gesture recognizing system using Convolution Neural Network (CNN) with modelling diverse motions, which includes the skin colour model enhancement, the hand stance is calibrated for boosting identification accuracy. A CNN based hand gesture recognition system is used in this project. CNN is subfield of a neural network research. There is no requirement to design sophisticated algorithms for extracting and learning visual information when making use of CNN to understand human gestures. Features are permitted with less dislocation across convolution and sub levels of a CNN. The primary axis of gesture is discovered to calibrate picture in order to lessen the influence of varied handpostures of a hand gesturing type on recognizing accuracy. Calibrated pictures aid a CNN's ability to learn and identify

accurately. Furthermore, the little condition and transitioning gestures on continuous motion has a significant impact on recognition accuracies, particularly the light condition. They use a Gaussian Mixture Model (GMM) to create robust skin structure in order to do better skin colour segmentation. Some transitive motions in a continuous gesture may be problematic to the CNN, thus they have used criteria to backward and foreward search of an image framing whose gesturing type is very much comparable to the transitive gesturing. This allows the system to filter out incorrectly identified gestures while maintaining the right gesture sequencing without affecting the actual meaning of the preceding motion.

Future scope and Implementation: This paper aims in developing a CNN-based hand gesture recognition system. The main feature of the system is that there is no requirement to build a model for each and every gesture using hand signs such as fingertips and contours. To have robust performance, they have applied a GMM for learning the skin model and segmenting the hand region for recognizing. The calibration of the hand posture is used for rotation and shifting the hand on image to a neutral position. A CNN has been trained for learning seven types of in this paper. They have performed four fold cross verification of the system in the trials, using 600 & 200 photos for training and testing, respectively, and the results revealed that the over all recognition rates of all seven gesture categories are around 99 percent.They have trained and evaluated hand photographs of the seven types of gesture from seven different people to test the proposed approach on numerous subjects. Average rate of recognition is about 95.90 percent. Using the given rules, the proposed system also produced good results for transitive gestures in continuous motion. According to the paper, higher-level semantic analysis must be applied to the existing current system in the future to improve recognizing capabilities for complicated human activities.

In the paper [6] have achieved the Sign language is a type of communique that use visible cues which include facial expressions, hand gestures, and frame moves to speak meaning. People who've hassle listening to or speak me may advantage significantly from signal language.

The translation of those gestures into phrases or alphabets of present officially spoken languages is called signal language recognition. Thus, the usage of an set of rules or a version to transform signal language into phrases can help bridge the distance among people with listening to or speech impairments and the relaxation of the world. Vision-primarily based totally hand gesture identity is a warm subject matter in pc imaginative and prescient and gadget studying proper now. It is a natural form of human contact, and many scholars are working to improve it. Therefore, the basic goal of gesture recognition research is to develop a system that can recognize and use specific human gestures, for example, to transmit information. The vision-based hand gesture interface requires fast, highly robust hand recognition and real-time gesture recognition. Hand gestures are a powerful mode of human communication that has multiple uses, and in this context sign language recognition, the communication technique used by the hearing impaired, is used. Hand gesture recognition for human-computer interaction is a hot topic in computer vision and machine learning research. The central goal is to develop a system that can recognize specific gestures and use them to send information and manipulate devices. It is a natural form of human contact, and many scholars are working to improve it. Therefore, the basic goal of gesture recognition research is to develop a system that can recognize and use specific human gestures, for example, to transmit information. The vision-based hand gesture interface requires fast, highly robust hand recognition and real-time gesture recognition. Hand gestures are a powerful mode of human communication that has multiple uses, and in this context sign language recognition, the communication technique used by the hearing impaired, is used. Hand gesture recognition for human-computer interaction is a hot topic in computer vision and machine learning research.

Methodology: This method is vision-based. All characters are rendered with bare hands, so no artificial gimmicks are needed for interaction.

A.DATASET GENERATION

Need to create an appropriate database of movements in sign language so that one can compare the photos collected during the interaction with this system.

Below are the steps used to create the dataset.

First, took about 800 pictures of each ASL icon for training purposes and about 200 pictures of each icon for testing purposes. First, capture each frame displayed by the machine's webcam. Determine the region of interest (ROI) for each frame, represented by a square with a blue border, as shown in the following image. I got the ROI which is RGB from the whole image.

B.GESTURE CLASSIFICATION

The approach used in this project is: This approach uses a two-tiered algorithm to predict the user's final icon.

Algorithm level 1: Apply Gaussian blur filters and thresholds to OpenCV frames to get the processed image after feature extraction.

2. The processed image is sent to the CNN model for prediction, and when the character is recognized over 50 frames, it is printed and used for word formation.

3. Space letters are used to represent spaces between words.

Algorithm layer 2: 1. Recognize multiple sets of symbols that produce equivalent results when recognized.
2. Then use a classifier specifically designed for these sets to distinguish them.

C.TRAINING AND TESTING

Convert the input RGB photo to grayscale and apply Gaussian blur to reduce unwanted noise. Scale the photo to 128x128 using the adaptive threshold to separate the hand from the background. After performing all the above steps, send the pre-processed input photo to the model for training and testing. Future Scope: A functional real- time visual-based recognition system for American Sign Language for the hearing impaired and dumb people was developed for the asl-alphabet. Using the dataset, have achieved a final accuracy of 92.0 percent. One can improve the predictions by creating a two-tiered algorithm that validates and predicts symbols that are similar to each other. In this way, virtually everyone can be identified.

The paper [7] is about customizing the sign language recognition dealing with cnn(convolutional neural network) using static signs. There are more than 35000 images of signs containing 100 static signs collected and experimented on 50 models. The main aim is to identify the Indian sign language. The technology is not high end with gloves and Kinect but uses Webcam to take images and extract features and classify. The purpose of using deep learning is because it outperforms machine learning method. The main technology used is CNN that consists of three layers which are convolutional layer, pooling layer and output layer. The methodology consists of 4 stages. They are acquisition of data, data processing, training the model and testing phase. The experimentation involves fine tuning the number of layers, filters and optimizer. The progress is checked in colored and grayed images. The results obtained is 99.9% accuracy and turns out best when the never of layers are reduced. The drawbacks are that the signs from video cannot be captured dynamically and requires more datasets for recognizing of the model. This paper proves to be more effective for Indian sign language recognition and can be further customized for application in mobile app based technology.

The above-mentioned papers bridged the gap only between ordinary people and hearing impaired population. Our proposed system also fills the gap between hearing impaired and blind population by enhancing the system with speech feature. Our system also stores the letters, forms words and then sentence which is much simpler to understand during a communication. Our proposed system outperformed by demonstrating its effectiveness on real and challenging sign language dataset.

III. Proposed Methodology

The new inventions and advanced Computer Vision based technology has made utilizing of sign language less complicated and much easier. The hand gesture is firstly filtered by using gray scale and Gaussian filter properties respectively to distinctly identify the gesture. Then the process goes through the classifier process for identification and filtering that predicts the motions made by hand. The result from the recognition of hand gesture will be converted to text and displayed onto the output. This project will be a great help to the deaf and dumb to communicate with ordinary people.

Our working model contains different modules related with each other and are integrated together by Python programming language as shown in Fig. 1.

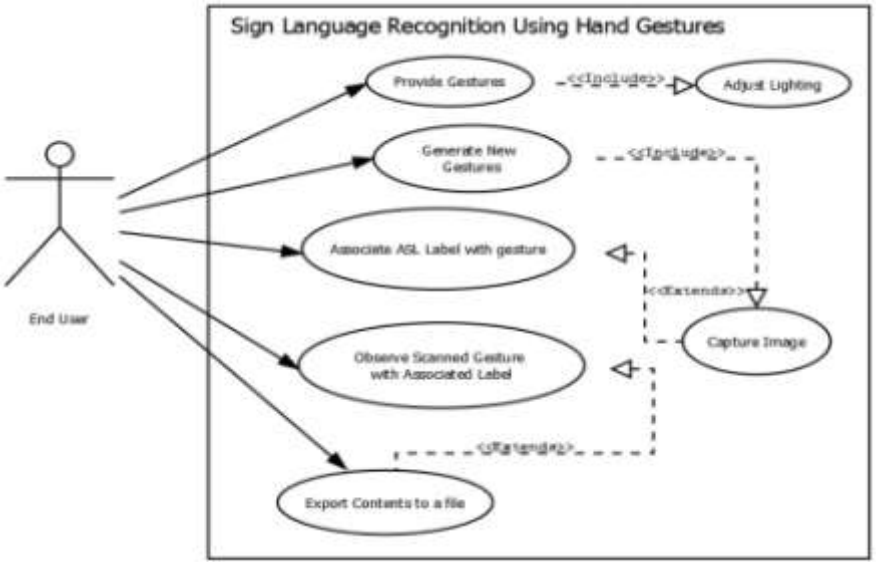

Fig. 1 . Use Case Diagram for Sign Language Recognition Using Hand Gestures

The proposed system works in two modes: Text mode and Calculator mode

TEXT MODE
DETECTS HAND GESTURES IN THE FRAME

Image-based hand motion recognition algorithms are used to recognize sign language. In order to express precise information, they employ a contemporaneous and specific mix of hand motions, hand shapes, and orientation.

RECOGNIZES THE LETTERS FROM HAND GESTURE

CNNs, or convolutional neural networks, are the most extensively used image categorization techniques. A gesture scanner will be placed in front of the end user, and the user will be required to make a hand motion. It should be possible for the user to view the linked label assigned to the hand motion based on the Pre-Processed module output.

ASSOCIATE SIGN LANGUAGE LABEL WITH GESTURE

A user will be providing a necessary hand gesture as an input to the system, and the space at the bottom of the screen will allow the user to enter the recognition word the corresponding gesture shown by the user. This customized gesture will be furthered saved for later use and identification for future use.

CONVERTS TO SPEECH

After assigning the label to the gestures, the predicted output will be converted to speech which enhances the system and increases efficiency.

CALCULATOR MODE
DETECTS HAND GESTURES IN THE FRAME

Image-based hand motion recognition algorithms are used to recognize sign language. In order to express precise information, they employ a contemporaneous and specific mix of hand motions, hand shapes, and orientation.

RECOGNIZES THE NUMBERS FROM HAND GESTURE

CNNs, or convolutional neural networks, are the most extensively used image categorization techniques. A gesture scanner will be placed in front of the end user, and the user will be required to make a hand motion. It should be possible for the user to view the linked label assigned to the hand motion based on the Pre-Processed module output.

ASSOCIATE SIGN LANGUAGE LABEL WITH GESTURE

A user will be providing a necessary hand gesture as an input to the system, and the space at the bottom of the screen will allow the user to enter the recognition word the corresponding gesture shown by the user. Thiscustomized gesture will be furthered saved for later use and identification for future use.

PERFORMS CALCULATION

A user inputs number and a corresponding operator (+, -, /, *, %, <<, >>, &, |) through hand gesture. Thesystem scans the gesture, performs the computation and the result will be displayed on the screen.

IV. Implementation

DATA PRE-PROCESSING

This module populates its binary pictures based on the item observed in front of the camera. That is, the item or any object must be in contrasting to the backdrop. As the backdrop will be in dark color mostly black theobject is expected to be in light color or preferably in white color. The following procedure for modules is givena collection of zero or one based on pixels regions.

SCAN SINGLE GESTURE

A gesture scanner will be placed in front of the end user, and the user will be required to make a hand motion. The person should view the appropriate characterization or tag allocated to each sign motion. The castration or tag is set according to the standard language to get preprocessed module output.

CREATING GESTURE

A gesture scanner will be placed in front of the end user, and the user will be required to make a hand motion. The person should view the appropriate characterization or tag allocated to each sign motion. The castration or tag is set according to the standard language to get preprocessed module output.

SENTENCE FORMATION

This system provides a delimiter until that 10 mm is found this system takes all other characters as input and this will be attached to the previous question this system then combines the letters to form words andfurther two form phrases and sentences. This helps individual to understand in a better way.

CONVERSION OF TEXT TO SPEECH

After assigning the label to the gestures, the predicted output will be converted to speech which enhances the system and increases efficiency. gTTS (Google Text-to-Speech) is a Python library, which is a very easy library that converts the text into audio.

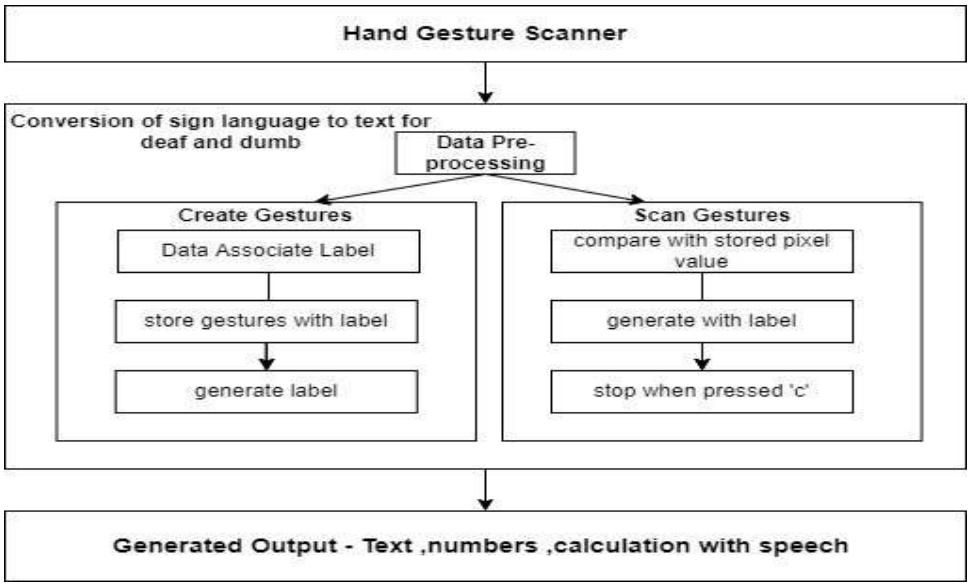

Fig. 2 . System Architecture for Sign Language Recognition Using Hand Gestures

V. Conclusion

In this project or application have tried to resolve some of critical problems faced by speech impairmentpeople. The project has aimed in finding out why they couldn't express themselves more freely. As the outcome many were not able to guess what deaf and mute people were trying to express. So this application serves the bridge gap between them. With the help of this application a person can quickly learn and adapt many gestures and their significance as per ASL standards. One can learn fast what alphabet is assigned to which gesture. Add- on to this custom gesture facility is also provided along with sentence formation. Users who know how gestures work do not have to be educated. One can quickly create gestures and the properly assigned characters will appear on the screen. A user does not need to be literate if they understand the expression; they may rapidly construct the motion and the proper allocated character appears on the screen. In terms of implementation, used the Tensor Flow framework in conjunction with the Keras API. In addition, PyQT5 is used to construct the entire front-end for user convenience. Appropriate user-friendly messages are prompted based on the user's activities, as well as which motion corresponds to which character window. Furthermore, an export to file module with TTS (Text-To- Speech) support is offered, which means that whatever phrase was generated, a user will be able to listen to it and then swiftly export while watching what gesture he/she did during sentence production.

References

[1] "Real-Time Translation of Indian Sign Language using LSTM" published in the year 2019 by the Author "E. Abhrehaam, A. Naik and A. Eqbaul

[2] Time Series Neural Network For Real-Time Sign Language Translation" by S Kumear, Sujay; Wagyel, Tenzin; Sebo, Varun; Shrinath, Ramamoorthy (2018). [IEEE 2018 17th IEEE

[3] "Sign Language Recognition based on adaptive HMMS with data augmentation" published in the year 2019 is by the Author "Dan Guo, Wengang Zhou, Meng Wang, and Houqiang Li

[4] Prof. Radha S. Shirbhate, Mr. Vedant D. Shinde, Ms. Sanam A. Metkari, Ms. Pooja U. Borkar, and Ms. Mayuri A. Khandge publishedan article named "Sign Language Recognition Using Machine Learning Algorithm" in IEEE.

[5] "Human Hand Gesture Recognition Using a Convolution Neural Network" published in the year July 2021 by the Author "Hsien I Lin, Ming-Hsing Hsu, and Wei Kai Chen

[6] A Review Paper on Sign Language Recognition for The Deaf and Dumb" published in the year 2021 by the Author R Rumana ,ReddygarSandhya Rani , Mrs. R. Premaa

[7] Deep learning-based sign language recognition system for static signs Ankita Wadhawan1 Parteek Kumar11 January 2020 Springer- Verlag London Ltd., part of Springer Nature 2020