FACULTY OF ENGINEERING AND TECHNOLOGY

# Department of Computer Science and Engineering

(Artificial Intelligence & Machine Learning)

Faculty of Engineering & Technology

Jain Global Campus, Kanakapura Taluk - 562112

Ramanagara District, Karnataka, India

**2022-2023**

**(5th Semester)**

**A business plan report on**

**Fundamentals of Innovation and Venture Development in Entrepreneurship – 3**

**(21PC3ED8)**

# "SIGN LANGUAGE TO TEXT"

**Submitted in partial fulfilment for the project centric learning (PCL) course**

## BACHELOR OF TECHNOLOGY

**IN**

## COMPUTER SCIENCE AND ENGINEERING

## (ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)

**Submitted by**


**Naman Shrimal, Priyal Saini, Syed Israr Hussain, Urjasvi Kurakula**

**21BTRCL069, 21BTRCL080, 21BTRCL109, 21BTRCL115**


**Under the guidance of**


**Ms. Sahana Shetty**

Department of Computer Science and Engineering

(Artificial Intelligence & Machine Learning)

Faculty of Engineering & Technology

JAIN (Deemed to-be University)

FACULTY OF ENGINEERING AND TECHNOLOGY

# Department of Computer Science and Engineering

(Artificial Intelligence & Machine Learning)

Faculty of Engineering & Technology

Jain Global Campus, Kanakapura Taluk - 562112

Ramanagara District, Karnataka, India

# CERTIFICATE

This is to certify that the idea titled **"SIGN LANGUAGE TO TEXT"** is carried out by **Naman Shrimal (21BTRCL069), Priyal Saini (21BTRCL080), Urjasvi Kurakula (21BTRCL115), Syed Israr Hussain (21BTRCL109),** third year bonafide students of Bachelor of Technology at the Faculty of Engineering & Technology, JAIN (Deemed-to-be University), Bangalore in partial fulfillment for the course Fundamentals of Innovation and Venture Development in Entrepreneurship – 3 in Bachelor of Technology in Computer Science and Engineering (Artificial Intelligence and Machine Learning) , during the year **2022-2023**.

**Ms. Sahana Shetty**
Assistant Professor
Dept. of CS&E,
(Artificial Intelligence and Machine Learning), Faculty of Engineering & Technology
JAIN (Deemed to-be University)

Date:

**Dr. Manujakshi B C,**
Head of the Department,
Computer Science and Engineering,
(Artificial Intelligence and Machine Learning), Faculty of Engineering & Technology
JAIN (Deemed to-be University)
Date:

# TABLE OF CONTENTS

# LIST OF FIGURES

# CHAPTER 1

## INTRODUCTION

This study and thesis explore the possibility of using artificial intelligence to enhance hearing and deaf people's communication. Like all other sign languages, Flemish sign language has its own vocabulary and syntax and is recognized by the government. Moreover, about 6000 Flemish people speak it as their mother tongue.

However, since so few people in the world are normally able to understand sign language, it can be challenging for hearing and deaf people to hold even basic conversations. In this way, the research tackles a problem that is extremely current and for which technology, like artificial intelligence, may be able to provide (part of) a solution.

This particular bachelor's thesis focuses on how Flemish Sign Language can be translated into written language using this technology. The question of whether it is possible to create a smart application that uses the device's camera to "look" at the deaf person's movements and then translate them into written or spoken words is being investigated. On the basis of the sign language dictionary, the hearing person can then speak or type a few sentences that are subsequently translated into the relevant sign language.

Making systems that can recognize particular gestures and utilize them to transmit information or control devices is one of its main objectives. However, hand postures are the static structure of the hand, whereas gestures are the dynamic movement of the hand, and gestures need to be described in the spatial and temporal domains. The two main methods used to recognize hand gestures are vision-based methods and data glove methods. The primary objective of this endeavor is to develop a vision-based system capable of real-time sign language recognition.A system based on vision is preferred because it offers a more straightforward and natural means of communication between a human and a machine.

The suggested method aims to comprehend some very basic signal language components and translate them to text and voice. A visible language is American Sign Language. The thoughts process linguistic information through the visual in addition to the signing. Form, positioning, motion, face emotions, and frame movements all play crucial roles in how information is conveyed. The entire USA does not speak sign language as their primary language. It Has its very own signal 6 language, and areas have vernaculars, like the numerous languages are spoken anywhere inside the globally speaking language, the detection rate by the ASL language as in compare to the grammatical accuracy is of 90 % percentage of institutions commonly use Indian sign language.

“A real-time portable sign language translation system” Paper demonstrates, a hand free demonstration of Taiwanese data language which uses the wireless system to process the data.To differentiate hand motion, they have inner sensors put into gloves to show the parameters as given by, posture, orientation, motion, defined of the hand in Taiwanese Sign Language could be recognize in no error. The hand gesture is considered by flex inner sensor and the palm size considered using the g sensor and the movement is considered using the gyroscope.Input signals would have to be consider for testing for the sign to be legal or not periodically. As the signal which was sampled can stay longer than the pre-set time, the legal gesture sent using phone via connectivity like Bluetooth for differentiating gestures and translates it. With the proposed architecture and algorithm, the accuracy for gesture recognition is quite satisfactory. As demonstrated the result get the accuracy of 94% with the concurrent architecture. Having a real-time sign language detector increases the efficiency of the community to able to in contact with people having disabilities like hearing and deaf society.[1]

“Deaf Mute Communication Interpreter” Paper: Have using machine learning algorithms presented the idea of a translation with skin colour tone to detect the ASL. They have made a skin colour segmentation that automatically depicts the colour and give the tune to it for further detection. They have used YCbCr spacing of colour as its vastly used in video template code and gives the efficient results of colour tone for human skin. Further they have taken the CbCr plane to distribute the skin tone colour. People from Different ethnicity have their tones different which is crafted in a model. Deep Learning methods and several machine learning algorithms are used to demonstrate translator to translate between parties.[2]

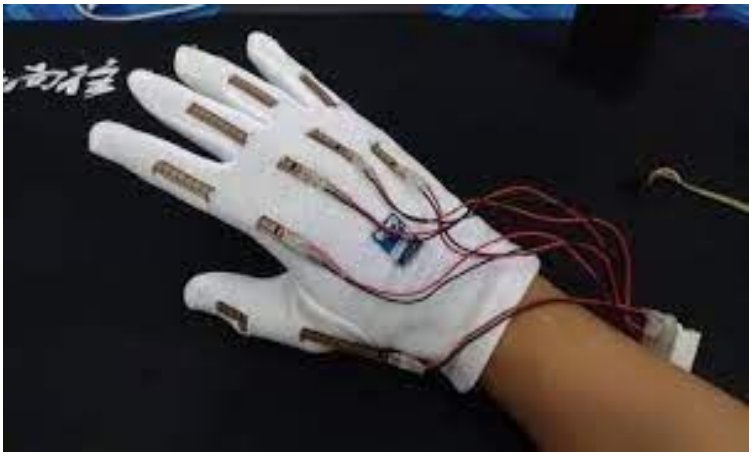

**Fig 1** **Hand Glove Machine**

“An Efficient Framework for Indian Sign Language Recognition Using Wavelet Transform Circuits and Systems” Paper: The authors applied a method of using a synthetic named animation making approach they have converted Malayalam language to Indian sign language. The intermediate representation in this method for sign language is being used by Ham-No-Sys. In this method the application accepts some sets of words, say, either one or more and forms It in animated portion. There is an interactive system which further converts the portion of words in to Ham-No-Sys designed structure. Its application pars everything that has been designed as it used by Kerala government to teach about sign language and subtle awareness.[3]

Having to communicate between deaf people and normal public has become a difficult task now days and to implement a such as the society lacks a good translator for it and having an app for it in our mobile phones is like having a dream at day.

“Hand Gesture Recognition Using PCA” Paper: The authors of this paper have proposed something great for the deaf community or hearing aid community by providing an app for the communication. But making an app for it is no simple task at it requires lot of efforts like memory utilization and a perfectly fined design to implement a such. What their application does is that they take a picture of a sign gesture and later converts is to a meaningful word. At first, they have compared the gesture using histogram that has been related to the sample test and moreover samples that are obliged to BRIEF to basically reduce the weight on the CPU and its time. They have explained a process on which on their app, its very easy to add up a gesture and store it in their database for further and expand detection set. So lastly, they came strong with having an app as a translator instead of several applications that are being used lately by users.[4]

“Hand gesture recognition system for dumb people” Paper : This paper is completely basedon a Spanish speaking language which converts the basic words into the Spanish language which is good for Spanish deaf people as it will provide them a stance to understand the sign language at pace as it will be converted in a Spanish language rather than in English which is popularly used as ASL. The device or an application that they have made for this comprises ofmany terms such as visible interface which is used by the deaf person to specify the sequence of sign data, a translator, which simply converts those series in Spanish based language in a formed series, and convertor to speech, which basically converts those entire bits into a meaningful sentence in Spanish, of course. They mainly focus on the interface which is designed visually for the impaired which has shown many ways to writing the sign language in a real time. There are some major techniques they have used in to translate as a final systemmodel. The mainly test data came from deaf people from cities like Madrid, Toledo and that itas a starting data which included measurements as an important form of data information. Its been the first designed Spanish corpus for a diverse research which target only to a specific domain. Containing more than 4000 sentences of Spanish language that is later being translatedinto LSE. With a several editions they have finally provided a fetched version of translator withdomain as renewal of Identification having records and having used pressure license.[5]

”International Journal of Science and Research” Paper: The authors have built a system which works in a continuous manner in which the sign language gesture series is provided to make a automate training set and providing the spots sign from the set from training. They haveproposed a system with instance learning as density matrix algorithm that supervises the sentence and figures out the compound sign gesture related to it with a supervision of noisy texts. The set at first that they had used to show the continuous data stream of words is further

taken as a training set for recognizing the gesture posture. They have experimented this set on a confined set of automated data that is used for training of them, identification for them and detection stored a subtle sign data to them. It has been stored around thirty sign language data that was extracted from the designed proposal. The Mexican Sign Language (LSM) is a language of the deaf Mexican network, which consists of a series of gestural symptoms and signs articulated thru palms and observed with facial expressions.[6]

“An Automated System for Indian Sign Language Recognition” Paper :The authors of this paper explained, the lack of automated structure to translate symptoms from LSM makes integration of listening to-impaired human beings to society extra difficult. This painting affords a totally new technique for LSM alphanumerical signs popularity based on 3D Haar-like features extracted from depth pictures captured by means of the Microsoft Kinect sensor. Features are processed with a boosting set of policies. To observe normal performance of our technique, we Identified a tough and speedy sign from letters and numbers, and in comparison, the results with the use of traditional 2D Haar-like capabilities. Our gadget is capable of recognize static LSM signs and signs with a higher accuracy percentage than the one obtained with extensively used 2D features.[7]

As it has been preferred for a society that having a sign language for hearing impaired and deaf people to communicate. As coming of several technologies in the past, it’s been kind of easy to have a translator which converts the sign language to the appropriate sentence and quite popular too.

“Hand Gesture Recognition for Sign Language Recognition” Paper : Its firstly based on an Arabic sign language which automates the process of being translated on to give a subtle way of communication and further they have shown that the scope of their project apars the usage and defined set of measurements. The application directly converts the Arabic sign language into a meaningful sentence by applying an automated Machine learning algorithm as they concluded.[8]

” Design Issue and Proposed Implementation of Communication Aid for Deaf & Dumb People” Paper : The authors have presented multiple experiments to design a statistical model for deaf people for the conversion to sign language from the speech set. They have further made the system that automates the speech recognition by ASR by the help of animated demonstration and translation statistical module for multiple sets of signs. As they went ahead,they used the following approaches for the translation process, i.e., state transducer and phrase defined system. As of evaluation certain figures type have been followed: WER, BLEU after that comes the NIST. This paper demonstrates the process that translates the speech by automation recognizer having all three mentioned configurations. The paper came up with theresult with finite type state transducer having the word error rate among the range of 28.21% and 29.27% for the output of ASR.[9]

The following” Sign-Language- Recognition-System” Paper depicts the usage of new and standardized model of communication system in which it basically targets the deaf people as that they have further explained. The system comprised of the following two scenarios like at first, Spanish speeches to Spanish sign translation that makes the usage of a translator to break up the words and have them converted into the stream set of signs which in along with each other makes a sentence and also depicts in an avatar form. Now the second scenario, the sign language that from a speech was generated then send to the generator, what this generator does

that it makes out those signs into the Spanish word which will make sense which basically converts the sign language into the Spanish based spoken language words. Now whatever the data previously was generated is comprises of words, now the last step is to convert the words into a sentence that is word to speech conversion. This paper consists of real example in city of Spain, Toledo that involved government personal and the deaf people. Further the paper describes the possible outcome and the scope of it in the future.[10]

"Automatic Sign Language Finger Spelling Using Convolution Neural Network: Analysis" Paper : It depicts the communication between deaf people and the other parts of society. It further depicts the movement of body, hands, fingers and other factors such as emotions on the face. The motion capture has been used to transmit the movements and for the sign language depiction. The dactyl-based modelling alphabet tech is developed using 3-d model. Having the efficiency on the Ukrainian sign language the realization is developed. This is based for the universal character and designed for further model.[11]

As times are improving and developing sign language emerges as one the best medium to communicate with the deaf community by having the supportive system to be a help to society.

The project they have demonstrated "Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning," Paper here are trying to make the communication easy by having the sign mainly having dynamic and static in ISL are being converted into the speech. A placed sensor of glove with sensor of flex help to design the orientation of hand and following actions. Using wireless transmission which converts it to the further bits of speech as the output. In this project they studied about LSTM networks which for long time formed dependencies. The result of this projects leads to a success rate of 98% accuracy which could able to identify the 26 gestures.[12]

Most of the research in this sector is conducted with a glove-based technique. Sensors like potentiometers, accelerometers, and other devices are mounted to each finger in the glove-based system. The corresponding alphabet is shown in accordance with what they read. A glove-based gesture recognition system created by Christopher Lee and Yangsheng Xu was able to recognise 14 of the hand alphabetic letters, learn new gestures, and update the model of each gesture in the system in real-time. Over time, sophisticated glove technologies like the Sayre Glove, Dexterous Hand Master, and Power Glove have been developed. The primary issue with this glove-based system is that everytime a new user logs in, it needs to be calibrated.

"Conversion of Malayalam text to Indian sign language using synthetic animation"Paper , we are doing this using image processing in our project. The key benefit of Our goal is to makeit so that it may be used with colours other than black background. With any background, it can be used. Likewise wearing in our system is not necessary to have colour bands. As a result, securities could confirm a person's identity based on "who she Rather than "what she possesses" and "what she could remember," use "what is".
[13]

Biometrics consists of two main classes:

- Physiological – It is associated with the body shape, includes all physical traits, iris, palm print, facial features, Fingerprints, etc.
- Behavioural – Related to the behavioural characteristics of a person. A characteristic widely used till today is signatures. Modern methods of behavioural studies are emerging such as keystroke dynamics and voice analysis.

“Deaf Mute Communication Interpreter” Paper : This study seeks to cover the various widely used techniques for biometrics system for deaf-mute communication interpretation. The two wide classifying the communication techniques that the People who are deaf-mutetypically use a wearable communication device.[14]

With regard to wearable communication There are three types of method: glove-based, keypad, and Touch-screen Handicom. All three of the aforementioned subdivided approaches require a variety of sensors, including an accelerometer and a text to speech conversion module, a suitable microcontroller, a touch-screen and a keypad. the requirement for an outside device to interpret the communication between a deaf-mute and a hearing person 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

" An Efficient Framework for Indian Sign Language Recognition Using Wavelet Transform " Paper : An example of a pattern recognition method is the ISLR system. This includes the crucial modules feature extraction and classification. using Discrete Wavelet Transform in tandem feature extraction based on (DWT) and closest neighbour classified is employed to understand sign language. The outcome of the experiment demonstrates how the suggested hand gesture recognition system works reaches maximum classification accuracy of 99.23% when using classifier using cosine distance.[15]

“Hand Gesture Recognition Using PCA“ Paper : The authors of this study provided a method for database-driven hand gesture detection based on thresholding and a skin colour model approach, as well as an efficient template matching, which may be used for human robotics applications and related applications. The segmentation of the hand region begins withthe use of the YCbCr colour space skin colour model. Thresholding is used in the following stage to distinguish between foreground and background. Principal Component Analysis (PCA) is then used to construct a template-based matching approach.[16]

“Hand Gesture Recognition System for the Dumb People”. 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: 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.

“Offline Signature Verification Using Surf Feature Extraction and Neural Networks Approach” .In this paper, an off-line neural network-based method for signature detection and verification is proposed, in which the user is shown a recorded image of the user's signature.

“Sign Pro-an Application Suite for Deaf and Dumb.The "Weakly Supervised Training of a Sign Language Recognition System Using Multiple Instance Learning Density Matrices," author described a sign language communication programme that enables the deaf and dumb to interact with the rest of the world. The system's real-time gesture to text conversion is its core component. Gesture extraction, gesture matching, and speech conversion are among the processing phases. Gesture matching methods include correlation-based matching and feature point matching. Other capabilities of the programme include text to gesture and voicing out of text. Using image segmentation and feature recognition algorithms, this study translated American sign language into audio and text. The system goes through a number of stages, including data collection, sensor, picture segmentation, feature extraction, etc.

This project is based on developing a desktop programme that records a user's ASL signing gestures and instantly converts them into related text and speech.

This project creates a machine learning model that can categorise the numerous sign language hand gestures. In this approach, a set of visual data is used to train machine learning classification algorithms.

The conversion of sign language to sentences involves recognizing and interpreting sign language gestures and translating them into complete sentences in the target language. This technology requires even more advanced algorithms that can accurately capture the complex syntax and grammar of sign language and translate it into grammatically correct sentences in the target language.

Capturing the complex syntax and grammar of sign language and translating it into grammatically correct sentences in the target language can be challenging. There may be multiple ways to express a single concept or idea in sign language, and identifying the most appropriate translation can be difficult. Accuracy and reliability of recognition technology is crucial, as any errors in interpretation can result in nonsensical or grammatically incorrect sentences.

## CHAPTER 2

### Explanation of designed system

Methodology

Our aim is to develop a system which interprets the sign language in English sentences. All existing systems [2, 3, 4, 5, 8] only focuses on recognizing the words which can be interpreted in wrong sentences so the ultimate goal is to convert the recognized continuous sign into proper English sentences. Proposed system details are as follows the system Model is mainly divided into two phases as follows: A. Sign language conversion into text i.e. words. B. Forming meaningful sentence of text using NLP techniques.

PHASE I:

Sign language conversion into text. Moreover work has been done in Sign language detection and conversion. This was done by using image processing techniques in combination with machine learning approach. In proposed work the sign language is detected and converted using following image processing techniques:

1. Input video

2. Framing (Key frame extraction)

3. Segmentation

4. Tracking

5. Feature extraction

6. Classification and Recognition

PHASE II:

Forming meaningful sentence of text using NLP techniques. The words that are recognized are then passed to the NLP engine where each word is given some description and this process is called tagging. POS tagging is divided into two stages tokenization and tagging. The raw text output gets tokenized and then each word is tagged associated part of speech.

Input video:
In this 2D camera upto 12 MP is used to capture the video. It provides the depth as well as skeleton information. 2D camera also reduces the cost of the process. This work does not address hand overlapping issue. The video taken by camera is further divided into frames. There are different type of video formats such as .avi, .mov, .wmv, .mng etc. We have done the work on .avi format video with maximum size of 8 MB.
Why .avi? : AVI is Audio Video Interleaved. We are using .avi video format because it supports any audio video coding formats. It uses RIFF i.e. Resource Interchange File Format (RIFF) which is a generic file container format that is used for storing data in tagged chunks.

- Grammar Design: To construct a valid sentence, we require a set of rules of the particular language these rules can be defined under Design grammar. It enables the formal representation of a vocabulary and rules that describe how designs can be synthesized.

- Parsing: The inputs to the parser are the parts of speech of the English language namely verb, noun, adjective, pronoun, conjunction, etc. Since stop words and articles are not included in the sign language, we will be inserting them into appropriate places using the lexical grammar tree for constructing a meaningful sentence. Consider the impaired user inputs. Here "Teacher give me apple and me share-with friends" The inputs to the parser are the parts of speech of English language viz.

- Grammar Check: After the sentence is formed we will verify the integrity and check where it contains any error. To accomplish this by using Language Tool in python which is an open-source grammar tool library that allows us to make, detect grammar to errors and spelling mistakes through a python script.

- DART: This type of technology could have many applications, such as automatically generating product descriptions from e-commerce product listings, or generating summaries of scientific research papers from structured data sources. The DART model is based on the GPT (Generative Pre-trained Transformer) architecture, which is a type of neural network that has been used for a wide range of natural language processing tasks. The DART model is trained on a large dataset of structured data records and their corresponding natural language descriptions, allowing it to learn how to generate accurate and informative descriptions from structured data.

- T5 (Text-to-Text Transfer Transformer): It is a state-of-the-art language model developed by Google's research team. It is a transformer-based model that was pre-trained on a massive amount of text data, and has achieved impressive results on a wide range of natural language processing tasks. It is trained on a wide range of tasks, it can be fine-tuned on a specific task with relatively few training examples.

Logistic regression:

One of the fundamental classification methods was where we started. We attempted to train a straightforward Logistic Regression model. The gradient descent algorithm and the sigmoid hypothesis function are used in logistic regression to train the data. We used our data to train the Sklearn logistic regression model. To prevent regularization, we set the penalty field of the parameter to none. Multiclass classification is our issue. The one-vs-rest approach is used in this model for multiclass categorization.

Moreover work has been done in Sign language detection and conversion. This was done by using image processing techniques in combination with machine learning approach. It uses a vision-based methodology. All the signs are represented with bare hands and so iteliminates the problem of using any artificial devices for interaction.

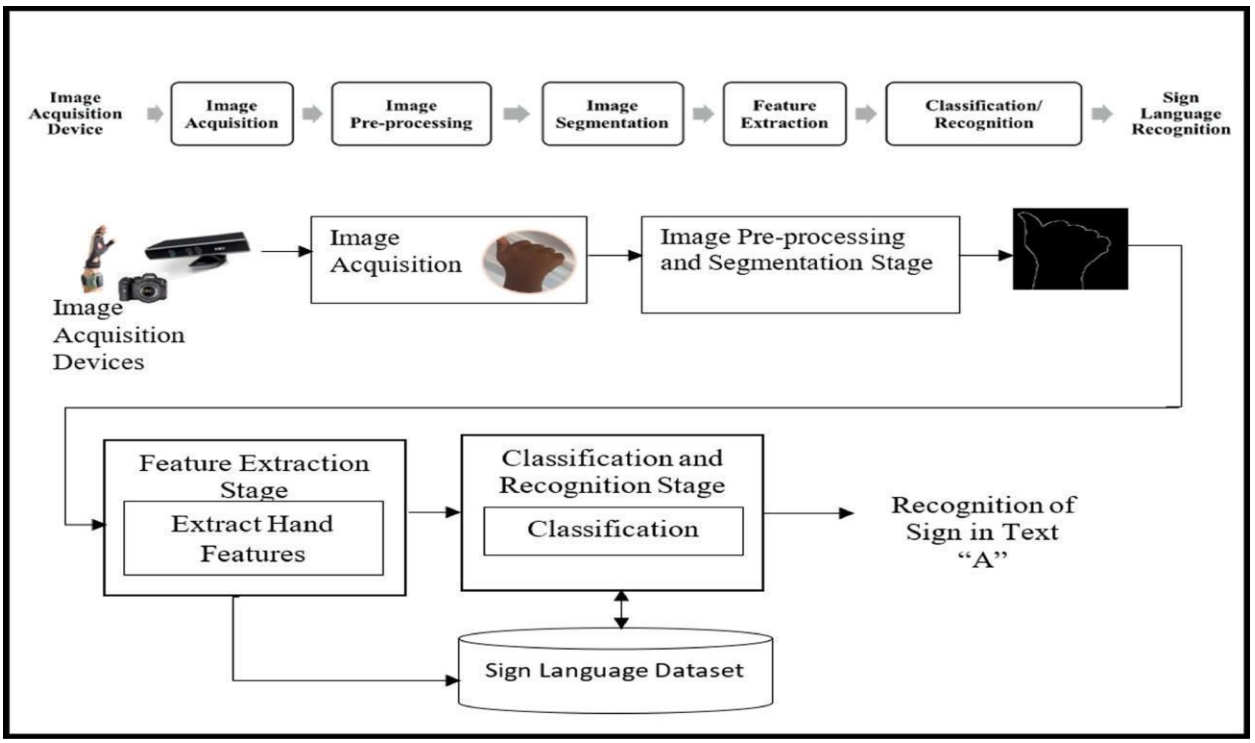


**Fig 2 System Design Flowchart**

## CHAPTER 3

## Implementation details

DATASET GENERATION

To compare the photographs taken while utilizing this technology for communication, a comprehensive database of sign language motions must be created. The procedures we used to produce our data set are listed below. To create our dataset, we utilized the Open Computer Vision (OpenCV) library. First, for training purposes, we took around 800 photographs of each ASL symbol, and for testing purposes, we took about 200 images of each symbol. We begin by taking a picture of each frame produced by our computer's webcam. As seen in the image below, each frame has a region of interest (ROI) that is indicated by a blue-bounded square. We retrieved our ROI, which is RGB, from the entire image and converted it into a greyscale image.

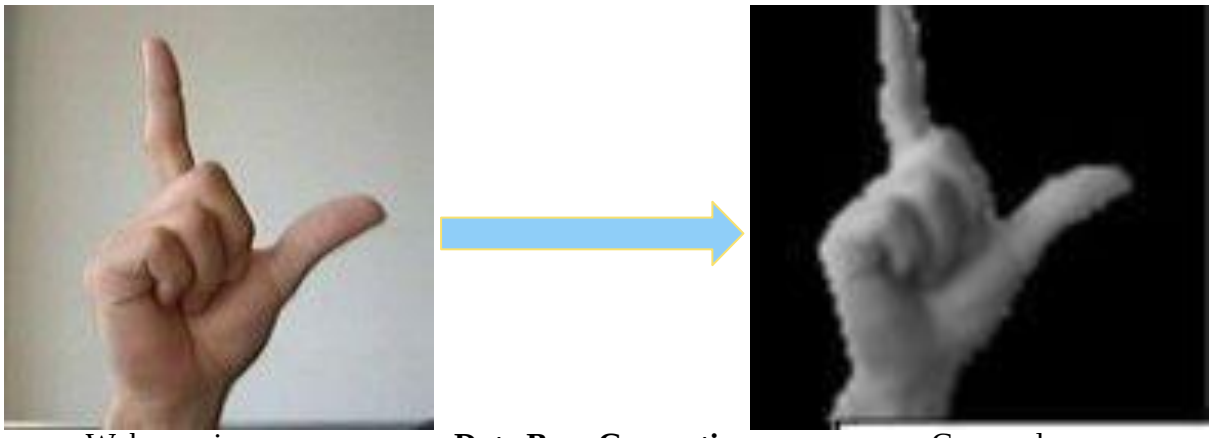


**Fig 3 Image processing of signs**

PROCESSING AND SEGMENTATION:

- Eliminate unwanted noise and enhance the quality of the image.
- Segment and extract the region of interest from the entire image.
- Thresholding, Background subtraction, skin-based and motion-based segmentation.

FEATURE EXTRACTION

- Method of reducing data dimensionality
- Transforms the input image region into feature vectors for recognition
- Centroid and skin color

All these pixels values can be presented directly to our model but this can result in challenges during modelling such as slower than expected training of the model. Instead, we believe it can be of great benefit in preparing the pixel values before doing any modeling such as standardization.

**Fig 4 Feature Extraction of hand gesture**

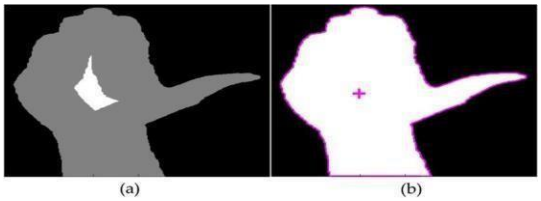


**Fig 5 Processing and Execution of hand gesture**

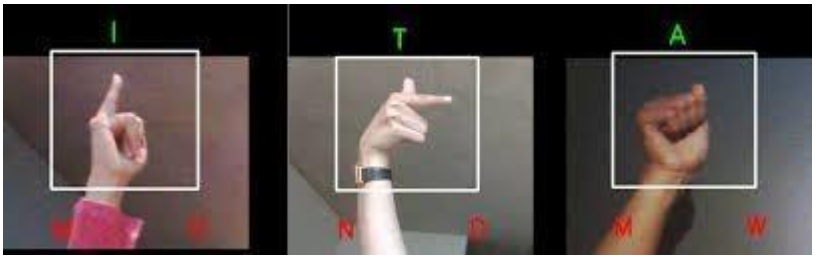


**Fig 6 Processing of hand gesture**

GESTURE CLASSIFICATION

Standardization of a dataset is a common requirement for many machine learning estimators: they might behave badly if the individual features do not more or less look like standard normally distributed data (example - Gaussian with 0 mean and unit variance). Standardization scales each input variable separately by subtracting the mean (called centering) and dividing by the standard deviation to shift the distribution to have a mean of zero and a standard deviation of one. Each pixel value is the feature of the dataset and since each pixel is importantfor an image so we need not do any feature selection/reduction. Our dataset does not have anymissing, NaN, noisy or inconsistent value so we feel we need not do any kind of Data-cleaning

The strategy that we employed for this project is

Our method predicts the user's final symbol using two levels of algorithm.

First Algorithm Layer:

After feature extraction, process the frame from an open camera with a gaussian blur filter and threshold to produce the final image.

A letter is printed and taken into consideration for creating a word if it is spotted for more than 50 frames in this processed image before it is sent to the CNN model for prediction.

Using the blank sign, the spaces between the words are taken into account.

Second algorithm layer:

We detect various sets of symbols which show similar results on getting detected. We then classify between those sets using classifiers made for those sets only.

Web cam Grayscale image Gaussian Blur filter

**Fig 7 Processing through grayscale**

Sentence generation for Indian Sign Language using NLP:

• The proposed system consists of two modules. They are, sentence generation using conventional NLP method, Sentence conversion using Bi-LSTM. These modules are combined to form a smart sentence conversion system for Indian Sign language.
• POS (part-of-speech) tagging is the process of identifying and assigning grammatical tags to each word in a sentence based on its syntactic category or part of speech. The part of speech of a word refers to the grammatical function that it serves within a sentence.
• Sentence Generation using conventional NLP modules involves the conventional NLP methods like POS tagging, grammar designing, parsing using LR parser. By tagging each word in the sentence by using the POS taggers. With the help of those tags the respective position of each word in the sentence gets analyzed.
• The conventional NLP method gave an accuracy of 90.6% and the Bi-LSTM model gave an improved accuracy of 95.6%

TRAINING AND TESTING:

We convert our input images into grayscale and apply gaussian blur to remove unnecessary noise. We feed the input images after pre- processing to our model for training and testing after applying all the operations mentioned above.

CHALLENGES FACED:

We experienced a lot of difficulties while working on the project. The dataset was the very first problem we ran into. Since working with merely square photos was much more practical, we wanted to handle raw images and that too in Keras . We opted to create our own dataset because we couldn't discover any existing ones for it. The second challenge was choosing a filter that we could use on our photos to extract the proper features, allowing us to subsequently input that image into the CNN model. We experimented with other filters, such as gaussian blur, binary threshold, canny edge detection, etc., but in the end, we settled on gaussian blur filter.

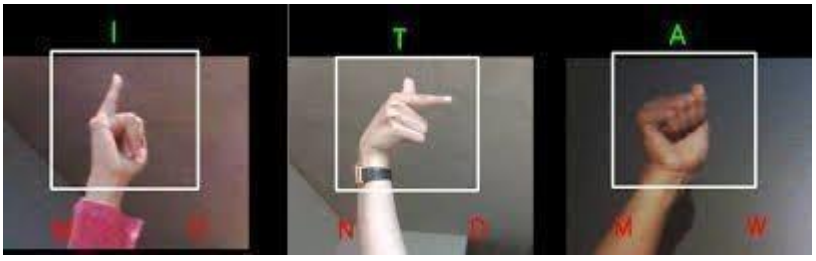


**Fig 8** **Hand gesture symbols**

Code snippets:

```
# Detect pose landmarks
with mp_pose.Pose(
    min_detection_confidence=0.5,
    min_tracking_confidence=0.5) as pose:
    results = pose.process(image)
```

```
# Draw landmarks on the image
    annotated_image = image.copy()
    mp_drawing.draw_landmarks(
        annotated_image, results.pose_landmarks, mp_pose.POSE_CONNECTIONS)
```

## Conclusion:

A system is developed that can recognize a set of Indian Sign Language gestures and convert them into meaningful text/speech using various Image Processing and Machine Learning/Deep Learning techniques. It makes a foundation for a scalable project that can be extended to capture the whole vocabulary of Indian Sign Language through manual and non-manual signs. Using a hybrid model gives the benefits of all the feature extraction techniques and gives substantial accuracy along with decreasing the computational time required. The sentences formed from the language model ensure the correctness and preciseness when compared to expected results. The suggested method aims to comprehend some very basic signal language components and translate them to text and voice. A visible language is American Sign Language. The thoughts process linguistic information through the visual in addition to the signing. Form, positioning, motion, face emotions, and frame movements all play crucial roles in how information is conveyed. The entire USA does not speak sign language as their primary language.
It Has its very own signal 6 language, and areas have vernaculars, like the numerous languages are spoken anywhere inside the globally speaking language, the detection rate by the ASL language as in compare to the grammatical accuracy is of 90 % percentage of institutions commonly use Indian sign language.

## Future Scope:

The system can be extended to include the knowledge of phonological and morphological information. Dependency Parsing and Text Similarity can be used if the input is a paragraph. The application can be extended to take speech as input and generate its corresponding gesture. We can also add more instances into the data set such that it covers all the spheres of communication and provides variations. A a personal assistant who actively interacts with people with disabilities can be developed by extending this system. From the keywords extracted, sentences can be generated in various languages like Hindi, Marathi, Gujarati, etc. with proper sentence formation schemes to ensure grammatically correct sentences.

REFERENCES

[1] L. Ku, W. Su, P. Yu and S. Wei, "A real-time portable sign language translation system," 2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS), Fort Collins, CO, 2015, pp. 1-4, doi: 10.1109/MWSCAS.2015.7282137.

[2] Sunitha K. A, Anitha Saraswathi.P, Aarthi.M, Jayapriya. K, Lingam Sunny, "Deaf Mute Communication Interpreter"- A Review, International Journal of Applied Engineering Research, Volume 11, pp 290-296, 2016.

[3] Mathavan Suresh Anand, Nagarajan Mohan Kumar, Angappan Kumaresan, "An Efficient Framework for Indian Sign Language Recognition Using Wavelet Transform Circuits and Systems", Volume 7, pp 1874- 1883, 2016.

[4] Mandeep Kaur Ahuja, Amardeep Singh, "Hand Gesture Recognition Using PCA", International Journal of Computer Science Engineering and Technology (IJCSET), Volume 5, Issue 7, pp. 267-27, July 2015.

[5] Sagar P.More, Prof. Abdul Sattar, "Hand gesture recognition system for dumb people",

[6] "International Journal of Science and Research" (IJSR)

[7] Chandandeep Kaur, Nivit Gill, "An Automated System for Indian Sign Language Recognition", International Journal of Advanced Research in Computer Science and Software Engineering.

[8] Pratibha Pandey, Vinay Jain, "Hand Gesture Recognition for Sign Language Recognition": A Review, International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 3, March 2015.

[9] Nakul Nagpal,Dr. Arun Mitra.,Dr. Pankaj Agrawal, "Design Issue and Proposed Implementation of Communication Aid for Deaf & Dumb People", International Journal on Recent and Innovation Trends in Computing and Communication ,Volume: 3 Issue: 5,pp- 147 149.

[10] S. Shirbhate1, Mr. Vedant D. Shinde2, Ms. Sanam A. Metkari3, Ms. Pooja U. Borkar4, Ms. Mayuri A. Khandge/"Sign-Language- Recognition-System".2020 IRJET Vol3 March,2020.

[11]M.V, Beena. "Automatic Sign Language Finger Spelling Using Convolution Neural Network: Analysis". (2017).

[12] S. Shahriar et al., "Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning," TENCON 2018 – 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 1168-1171, doi: 10.1109/TENCON.2018.8650524.

[13] M. S. Nair, A. P. Nimitha and S. M. Idicula, "Conversion of Malayalam text to Indian sign language using synthetic animation," 2016 International Conference on Next Generation Intelligent Systems (ICNGIS), Kottayam, 2016, pp. 1-4, doi: 10.1109/ICNGIS.2016.7854002.

[14] M. Mahesh, A. Jayaprakash and M. Geetha, "Sign language translator for mobile platforms," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 1176-1181, doi: 10.1109/ICACCI.2017.8126001.

[15] S. S Kumar, T. Wangyal, V. Saboo and R. Srinath, "Time Series Neural Networks for Real Time Sign Language Translation," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, 2018, pp. 243-248, doi: 10.1109/ICMLA.2018.00043.

[16] D. Kelly, J. Mc Donald and C. Markham, "Weakly Supervised Training of a Sign Language Recognition System Using Multiple Instance Learning Density Matrices," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 41, no. 2, pp. 526-541, April 2011, doi: 10.1109/TSMCB.2010.2065802.

APPENDIX A: Sample source code

```
import cv2
from cvzone.HandTrackingModule import HandDetector
from cvzone.ClassificationModule import Classifier
import numpy as np
import math
import time
import tensorflow as tf

cap = cv2.VideoCapture(0)

detector = HandDetector(maxHands=1)
classifier = Classifier("Model/keras_model.h5", "Model/labels.txt")

offset = 20
imgSize = 300

labels = ["call", "Excellent", "father", "hello", "help", "iloveyou", "me", "need",
          "no", "peace", "thank you", "thumbsdown", "thumbsup", "what", "yes",
"you"]

ch = 0
cnt = 0
sentence = ""
fsentence = ""

while True:
    success, img = cap.read()
    imgOutput = img.copy()
    hands, img = detector.findHands(img)

    if hands:
        hand = hands[0]
        x, y, w, h = hand['bbox']

        imgWhite = np.ones((imgSize, imgSize, 3), np.uint8) * 255
        imgCrop = img[y - offset:y + h + offset, x - offset:x + w + offset]

        if imgCrop.size == 0:
            print("Error: imgCrop is empty.")
            continue  # Skip the current iteration if imgCrop is empty

        imgCropShape = imgCrop.shape
        aspectRatio = h / w

        if aspectRatio > 1:
            k = imgSize / h
            wCal = math.ceil(k * w)
```

```
            imgResize = cv2.resize(imgCrop, (wCal, imgSize))
            imgResizeShape = imgResize.shape
            wGap = math.ceil((imgSize - wCal) / 2)
            imgWhite[:, wGap:wCal + wGap] = imgResize
            prediction, index = classifier.getPrediction(imgWhite, draw=False)

        else:
            k = imgSize / w
            hCal = math.ceil(k * h)
            imgResize = cv2.resize(imgCrop, (imgSize, hCal))
            imgResizeShape = imgResize.shape
            hGap = math.ceil((imgSize - hCal) / 2)
            imgWhite[hGap:hCal + hGap, :] = imgResize
            prediction, index = classifier.getPrediction(imgWhite, draw=False)

        if ch == index:
            cnt += 1
        else:
            ch = index
            cnt = 0

        if cnt == 10:
            sentence += labels[index]
            sentence += " "
            if sentence == "hello thumbsup thumbsdown ":
                fsentence = "Get Lost"
            cnt = 0

        cv2.rectangle(imgOutput, (x - offset, y - offset-50),
                      (x - offset+90, y - offset-50+50), (255, 0, 255), cv2.FILLED)
        cv2.putText(imgOutput, labels[index], (x, y - 26),
                    cv2.FONT_HERSHEY_COMPLEX, 1.7, (255, 255, 255), 2)

        cv2.rectangle(imgOutput, (x-offset, y-offset),
                      (x + w+offset, y + h+offset), (255, 0, 255), 4)

        cv2.imshow("ImageCrop", imgCrop)
        cv2.imshow("ImageWhite", imgWhite)

    cv2.putText(imgOutput, sentence, (10, 50),
                cv2.FONT_HERSHEY_COMPLEX, 0.8, (255, 255, 255), 2)
    cv2.putText(imgOutput, fsentence, (10, 300),
                cv2.FONT_HERSHEY_COMPLEX, 1, (255, 255, 255), 2)

    cv2.imshow("Image", imgOutput)
    cv2.waitKey(1)
```

APPENDIX B: Screenshot

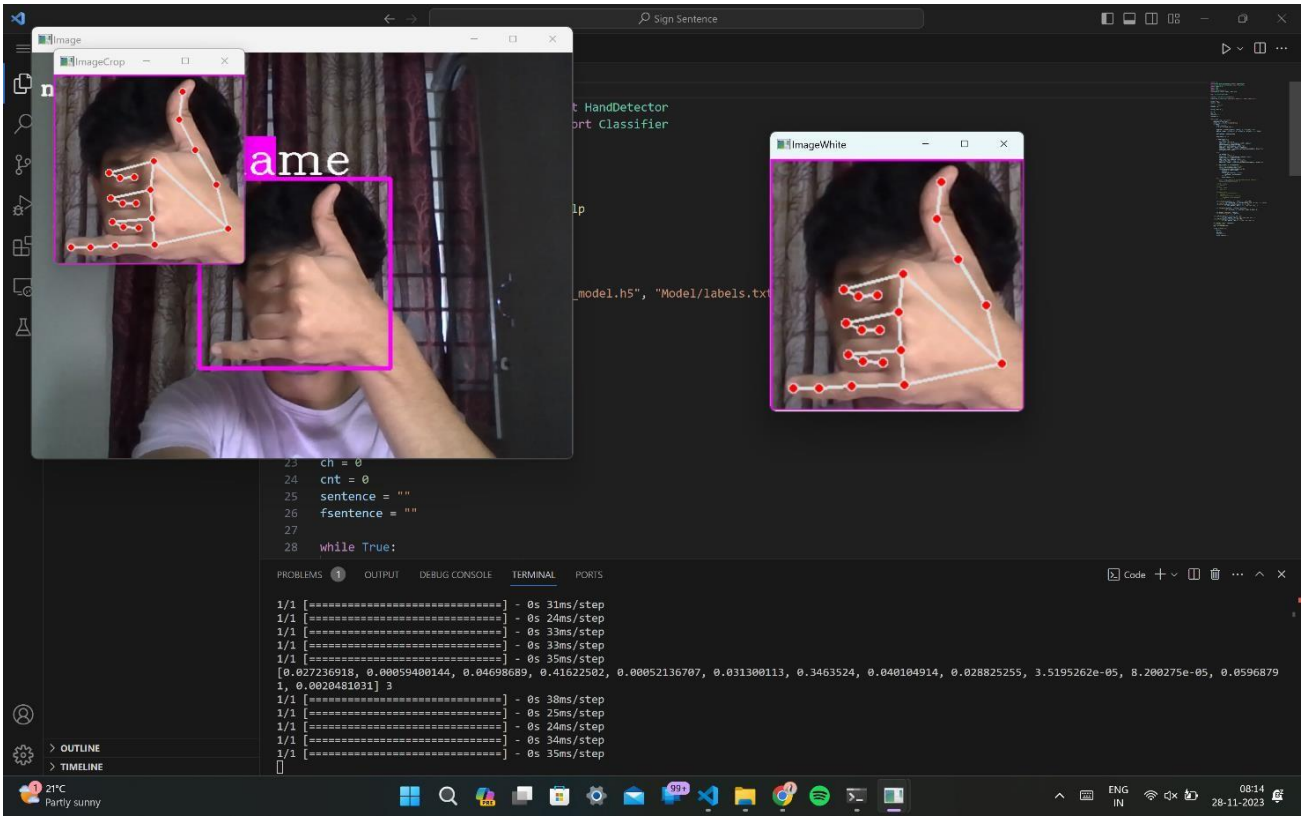


**Fig 9. Screenshot of implementation**