# INDIAN SIGN LANGUAGE INTERPRETER

**A PROJECT REPORT**

***Submitted by***

**MUHAMMED JAZIL (2116210701168)**

***in partial fulfillment for the award of the degree***

***of***

**BACHELOR OF ENGINEERING**

***in***

**COMPUTER SCIENCE AND ENGINEERING**



**RAJALAKSHMI ENGINEERING COLLEGE**

**ANNA UNIVERSITY, CHENNAI**

**MAY 2024**

# RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

## BONAFIDE CERTIFICATE

Certified that this Thesis titled **"INDIAN SIGN LANGUAGE INTERPRETER**" is the bonafide work of "**MUHAMMED JAZIL JAYAFAR (2116210701168)"** who carried out the work under my supervision. Certified furtherthat to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred onan earlier occasion on this or any other candidate.

**SIGNATURE**

Dr . S Senthil Pandi M.E.,Ph.D.,

**PROJECT COORDINATOR**

Professor

Department of Computer Science and Engineering

Rajalakshmi Engineering College

Chennai - 602 105

Submitted to Project Viva-Voce Examination held on_____

**Internal Examiner** **External Examiner**

# ACKNOWLEDGMENT

First, we thank the almighty god for the successful completion of the project. Our sincere thanks to our chairman **Mr. S. Meganathan B.E., F.I.E.,** for his sincere endeavor in educating us in his premier institution. We would like to express our deep gratitude to our beloved Chairperson **Dr. Thangam Meganathan Ph.D.,** for her enthusiastic motivation which inspired us a lot in completing this project and Vice Chairman **Mr. Abhay Shankar Meganathan B.E., M.S.,** for providing us with the requisite infrastructure.

We also express our sincere gratitude to our college Principal, **Dr. S. N. Murugesan M.E., PhD.,** and **Dr. P. KUMAR M.E., PhD, Director computing and information science , and Head Of Department of Computer Science and Engineering** and our project coordinator **Dr. K.Ananthajothi M.E.,Ph.D.,** for her encouragement and guiding us throughout the project towards successful completion of this project and to our parents, friends, all faculty members and supporting staffs for their direct and indirect involvement in successful completion of the project for their encouragement and support.

**MUHAMMED JAZIL JAYAFAR**

# INDIAN SIGN LANGUAGE INTERPRETER

**ABSTRACT**- Sign language serves as a visual means of conveying messages through hand movements, alterations in hand shape, and tracking gestures, primarily for individuals facing hearing and language impairments. Given the challenges encountered by those with speech impairments, this system provides a tool to enhance communication and bridge gaps. The primary objective of this research effort is to implement a tool that enhances communication for those with difficulties with speech. The objective is to devise a system enabling individuals with speech impairments to engage in two-way conversations, even in noisy environments. LSTM networks were studied and used to sort gesture data because they understand long-term relationships. The technology combines computer vision skills to incorporate real-time sign language motions of those with speech impairments. Edge detection methods proficient in text and voice recognition are utilized for hand recognition, complemented by deep neural networks for motion recognition. By gathering datasets from individuals' videos and utilizing comprehensive key points to identify poses, facial expressions, and hand gestures, the model undergoes training. Ultimately, the system displays appropriate hand movements and converts speech to text. An impressive 89% success rate for this model shows how well LSTM-based neural networks can be used for sign language translation.

**Keywords**—Sign Language Interpreter, NLP, CNN, LSTM

## INTRODUCTION

It is important to note that machine learning (ML) is a key part of artificial intelligence (AI) since its inception in 1959 by Arthur Samuel, as outlined in "Machine Learning" (2019). Over the past decades, ML has significantly influenced a wide array of applications globally. Within the ML domain, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are prominent architectures, with CNNs often categorized under Deep Learning (DL), as noted by Alom et al. (n.d.)[1]. These technologies find extensive usage in various AI applications such as image recognition, speech understanding, medical diagnostics, classification, and forecasting, as highlighted in "Convolutional Neural Network" (2019). Additionally, the paper delves into a system aimed at recognizing gestures,

a form of sign language commonly utilized by individuals with hearing impairments or speech disabilities.

Sign language is the main way that people who are deaf or have trouble speaking can communicate.. It entails using manual gestures to convey ideas and feelings. Each country, including those with speech and hearing impairments, has developed its unique sign language, varying across nations. Indian Sign Language (ISL) stands as India's indigenous sign language, comprising approximately 350 signs, with its alphabet containing 26 letters, differing from the conventional English alphabet (Stone, n.d).

Over the preceding two decades, the technological landscape has witnessed the emergence of numerous gesture recognition systems within academic discourse. These encompass a diverse array of methodologies, including Glove-based systems, Some examples of these methodologies are Glove-based systems, k-NN algorithm application, ELM, HMM, DTW, Finite State Machine, Fingertip Search, Hu Moments, Eigenvalue weighted Euclidean distance, and Grid-Based Feature Extraction for feature extraction.

According to data from the 2018 Census of Population and Housing in India, the population included 569,910 individuals with concurrent hearing and speech impairments (India, 2018). However, the needs of the deaf and speech-impaired community in India often go unnoticed by the general public, mainly due to language barriers. Consequently, these individuals are often marginalised, hindering their ability to lead ordinary lives. Language disparities pose significant challenges to effective communication among the deaf and speech-impaired, limiting their access to necessities and desires compared to those without such impediments. Moreover, the limited availability of facilities catering to the disabled in India presents additional obstacles for service providers seeking to address their needs comprehensively. The linguistic gap exacerbates feelings of isolation among the deaf and speech-impaired, increasing their susceptibility to depression, as documented by Flexer (n.d.)[4]. Research has underscored the heightened risk of psychological distress, including depression and anxiety, faced by this demographic, which is twice that of the general population. This paper advocates for a solution to mitigate this disparity and ensure equitable opportunities for this community.

An improvement in accessibility for those with speech impairments is the primary goal, which is to promote communication between sign language users and those who use written or spoken language. This framework aims to obviate the necessity

for an intermediary interpreter by directly translating sign language gestures into speech or text output. Such a system endeavors to create a user-friendly environment by providing seamless communication with minimal intermediation. The role of interpreters in bridging communication gaps between hearing individuals and those with hearing impairments demands a high level of linguistic proficiency, adeptness in comprehension of spoken language, global awareness, adherence to ethical standards, and demonstration of professionalism. Presently, interpreters face limitations in their ability to comprehensively translate and interpret, necessitating their involvement in various conversational contexts. Although interpreter services are commonly associated with individuals with hearing impairments, the ubiquity of the need for interpretation underscores the universal desire for effective communication. The primary goal of this endeavor is to obviate the requirement for interpreter assistance in routine interactions by developing an accessible application capable of real-time interpretation of sign language captured via camera input, thereby catering to diverse user preferences.

In the present-day context, individuals often find themselves engrossed in various activities, leaving limited time for the acquisition of sign languages. The pursuit of this initiative stands to enhance communication between individuals and those with disabilities. Unlike conventional approaches reliant on cumbersome systems, this study presents an advancement by employing an LSTM (Long Short-Term Memory) model to facilitate real-time translation of videos into text, while also maintaining a lightweight framework. This collaborative application targets individuals with hearing impairments, encompassing diverse modalities of sign language and gestural communication.

## LITERATURE SURVEY

SVBiComm, developed by Reda et al., is a computer vision system devised to interpret gesture language. The process involves transforming images into textual representations, which are then converted into audible speech. SVBiComm functionalities encompass bidirectional conversion of speech to video and vice versa. Text undergoes image processing techniques for processing, followed by its assignment to a 3D avatar synchronised with the voice assigning tasks to a voice-synchronized 3D avatar.

"Proposed SSVM Classifier and Hand sign acceptance method" is the name of the study that is cited as following:. Capturing the image, segmenting the skin tone,

removing the background, detecting edges accurately, and extracting PCA features are the eight stages outlined by Kumar et al. Image classification using data collection, examination, and support vector machine methods is what this procedure is all about. The classifiers are analysed in MATLAB with an emphasis on image colour analysis after being recorded by the camera. The RGB model transforms color images into binary format, converting skin tones into black-and-white pixels. Further processing involves cropping images using the bwareaopen command and applying efficient edge detection methods. The acute edge detection tool supports vector-based training, testing, and classification. The SVM detects various movements, determines image categorization, and adapts plans accordingly. Experimental results showcase a 94% accuracy level.

In their investigation, Tao et al. (cited as [7]) proposed a methodology for the recognition of American Sign Language (ASL) alphabets. Their approach involved the utilization of Convolutional Neural Networks (CNNs) in conjunction with multiview expansion and fusion techniques for the analysis of high-resolution images acquired from Microsoft Kinect. The multiview extension technique augmented the dataset by capturing 3D information from various perspectives, enhancing the model's ability to account for perspective variations inherent in real-world scenarios. Comparative analysis revealed superior performance of this approach over conventional imaging methods, particularly in accurately reproducing perspective variations present in real-world environments. Challenges arising from variations in perspective and finger movements across different sign classes were addressed through a process of combining inferences. The culmination of data from multiple perspectives resulted in improved model performance, as evidenced by experimental results demonstrating recognition accuracies of 100%.

In their research, Ananth Rao et al. (Reference [8]) devised a novel approach for the continuous recognition of Indian individuals using selfie videos. This unique video format integrates gesture language and relies exclusively on the computational capabilities of smartphones. Classification, prefiltering, pattern identification, and Sobel operations are just a few of the many uses for Gaussian filtering, which makes use of framework subtraction and a modulable block threshold. Discrete cosine transformation is used to classify the degree of detail of the hand and head contours. Execution speed is improved by removing unnecessary parts and concentrating on the essentials. Standardised area metrics based on the work of Euclidean and Mahalanobis are used to classify sign features. Nevertheless, the most difficult part was correctly identifying signers' hand shapes. In a perfect world, signers would only

need to use one hand, but in practice, things may become messy when the signer moves the selfie stick or when the camera shakes.

In their study, ElBadawy et al. [9] used a 3D CNN to correctly recognise and categorise 25 distinct signs from the ArSL lexicon. The data used by the recognition algorithm came from depth maps. The system takes in normalized depth as input and refers to it as a video stream. Initially, the video data is inputted and segmented into individual frames, which are then downsampled and ranked using a scoring system. The system's architecture discerns spatial and temporal components from the input, categorizing them through the softmax layer. Based on the results, the three-dimensional deep architecture is quite effectiveThe identified gestures are then shown as text in both English and Hindi and sounded out in both languages after the image processing is complete. The next steps include feature extraction and classification using KNN using characteristics extracted using the 7Hu moment technique. The precision rate was 80%. The integration of spoken English into MATLAB is quite effective. Given the time constraints of video conferencing, the experimental results show that PNN integration improves the sorting rate, which could have applications in improved communication with the hearing impaired.

Tanuji Bhola and his coworkers created a system that lets two people communicate using sign language in real time. The system uses computer vision, picture processing, and machine learning techniques. To enhance performance, methods such as median blurring, edge detection, hand identification, and skin color segmentation are applied to the image dataset. The system achieves an impressive precision of 99% for forecasting 17,500 simulated images within a small 14-second timeframe using a CNN model that was trained on a large dataset with 40 classes.

In their study, Karen Das et al. [11] introduced a technique aimed at interpreting Indian Sign Language through real-time video analysis. The system comprises three consecutive phases. The initial step involves preprocessing, incorporating skin filtering and histogram matching. After that, we look at the eigenvalues and eigenvectors to do feature extraction. Lastly, the eigenvalue-weighted distance determined by Euclid serves in the analysis phase. Included in the study's dataset were their data.

A framework for Arabic gesture recognition using a Convolutional Neural Network, also referred modelled after LeNet-5 was proposed by Salma Hayani and coworkers. There were 7869 pictures of Arabic letters and numbers in the collection. Studies used training data in varying proportions, 80%. The system achieved a 90% accuracy

rate while utilising the training dataset at an 80% rate. There was also a comparison of the machine's effectiveness with two other machine learning methods, k-nearest neighbour (KNN) and Support Vector Machine (SVM). The purpose of developing this model was to address challenges related to video and picture recognition.

## SYSTEM DESIGN

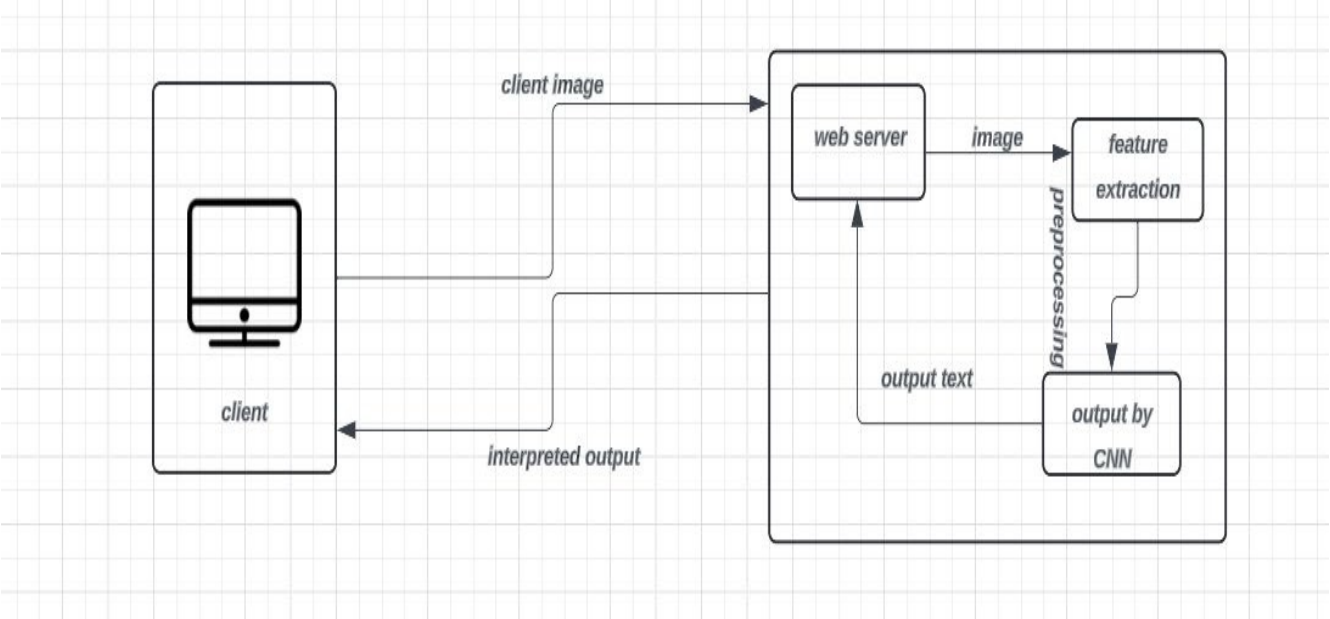



Fig.1 System architetcture

I. Packages Installing

Machine learning algorithms include convolutional neural networks. Having nodes or neurons linked via weighted linkages that provide an output corresponding to the input is a hallmark of these networks, which are reminiscent of artificial neural networks. Convolutional networks excel at visual classification tasks, such image processing, which is the key distinction. There is an output layer in a typical neural network that provides the categorization result, and a hidden layer that links to the input layer before it. On the other hand, typical neural networks aren't designed to process massive datasets. So, convolutional neural networks operate better with many images.

## II. Dataset

To generate our dataset, we'll capture live video from the webcam. This is achieved by using the cv2.VideoCapture class to create a VideoCapture object in Python. With VideoCapture, we specify either the file name or device index of the video source. The device index, represented by an integer, determines which camera to use: 0 for the primary camera, 1 for the secondary, and so on. Subsequently, we systematically capture each frame of the video. Using the OpenCV library, we extract video frames and apply the MediaPipe Holistic algorithm to detect key points corresponding to hand gestures, facial features, and body posture. The aim is to extract and store significant features from 30 sequences, each comprising 30 frames, for every action intended for inclusion in our model. The numpy file format is used to hold these recognised prominent features. A collection of numpy files with the parameters of the main elements and dimensions 30*30 is used to represent each activity.

Therefore, if the model requires incorporating n actions, the total data collected would amount to n * 30 * 30 numpy files. It's important to note that key points and values pertain to the prominent landmarks of the hand, face, and stance, as defined by MediaPipe Holistic. The MediaPipe framework enables the creation of an open-source, cross-platform media pipeline tailored for practical machine learning applications. This technology paves the way for the creation of advanced machine learning models, which can detect and track objects, recognise faces, and more. With MediaPipe as your guide, incorporating models into systems on all sorts of platforms is a breeze. This frees up developers to focus on model research instead of the nitty-gritty of system development.

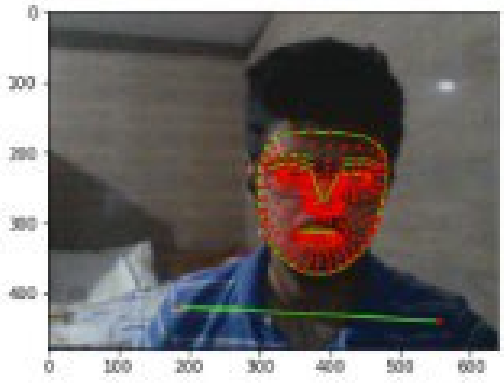



Fig.2 Landmark

The Holistic MediaPipe Framework, developed by Google, is a specialized tool designed for implementing machine learning solutions. It stands out for its open-source nature and compatibility across different platforms, making it widely accessible and adaptable MediaPipe is a dependable option for finger and hand tracking because it provides models have been trained for a variety of applications, including facial recognition, identification of objects, position estimation, and hand tracking. From a single frame, it uses machine learning to deduce the locations of 21–33 D local hand landmarks.. Unlike many existing methods that focus on desktop applications, our solution delivers real-time performance on mobile devices, ensuring scalability. By making this tool available to a wide community of researchers and developers, our aim is to encourage the creation of innovative applications, thus fostering the advancement of new research methodologies and applications.

## III. Preprocessing

We've developed a label map to assign specific actions, like 'hello' with a value of 1, 'thanks' with a value of 2, 'I love you' also with a value of 2, and so forth. Our data, which currently contains over 1500 values per frame, will be restructured into a single extensive array. This array will consist of 90 smaller arrays, each representing a sequence of 30 frames. Within these sequences, each frame will contain 1662 values. The target dimensions for the final array 'x'.

## IV. Train and Test Split

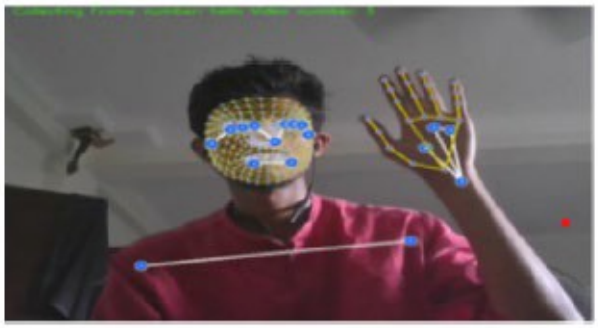
Use an experiment size of 0.05 and the train_test() function to divide the data Layers 1 and 3 of the three LSTM layers should have 64 nodes each, while Layer 2 should have 128. Layer 3's return sequences are set to negative, and Stages 1 and 2's return sequences to true so that all three layers use the ReLU activation function. A trio of Dense layers should be added after the LSTM layers. Both Layer 4 and Layer 5 will use the'relu' activation function; the former will have 64 nodes and the latter will have 32. Lastly, the'softmax' activation algorithm will be employed in the output layer. Once the model is configured, you can start training it by using it.modelling and compilation.execute fit() with the specified parameters. You can use TensorBoard to see how well the model is doing.

## V. Output

It is expected that our target accuracy would surpass 80%. Keep in mind that we ran into a similar problem when training the model, so it may be required to gather data again if the category accuracy goes below 0% when calculating the model's accuracy. Next, you can use OpenCV to connect the model to your video input—preferably a camera—so you can translate gestures into text in real-time after you've evaluated it.

# RESULT

Fig.4 Output (image depicting hi)



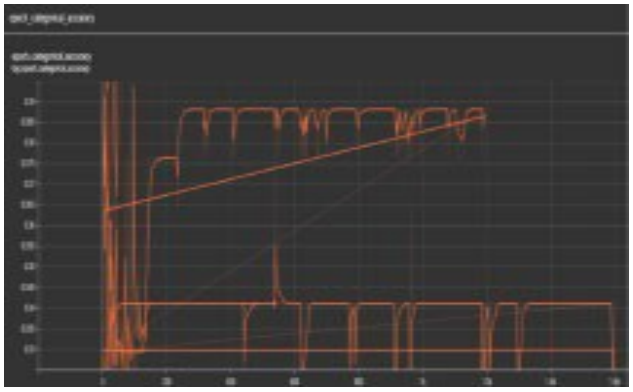


Fig.5 Training and testing accuracy.

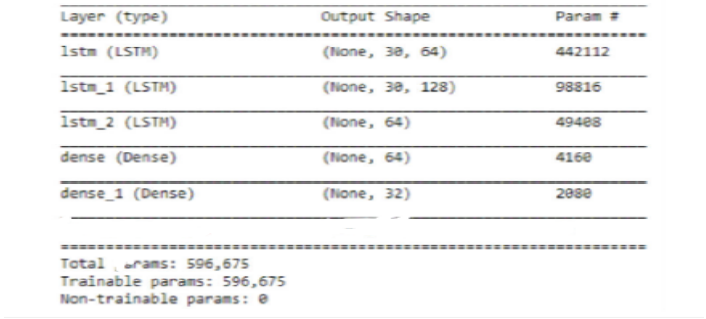



Fig.6 CNN model

## CONCLUSION AND FUTURE ENHANCMENET

This study aims to tackle the challenges faced by individuals with speech impairments. By utilising extraction of features methods, the Mediapipe library helps to streamline and optimise the model architecture, which in turn reduces computing overhead during classification. These techniques outperform more conventional classifiers like KNN and SVM. With a prediction accuracy exceeding 90%, this model serves as a solid foundation for further improvements in real-world applications.

Accuracy holds significant importance in Sign Language Detection. Presently, we have achieved a 90% accuracy rate utilizing LSTM. However, there remains a 10% chance of encountering variations in outcomes with this module. Moving forward, our goal is to enhance the accuracy of this component to nearly 100% by exploring alternative algorithms and techniques, enabling the system to make more precise predictions.

## REFRENCES

[2] The paper by Kumar et al. (2018) uses SSVM classifiers and hand gestures to recognise sign language. In: Intelligent Circuits and Systems International Conference (ICICS). Hagwara.

[3] Traditional neural networks enhanced with multiview and inference fusion for American Sign Language alphabets recognition [3] Tao W, Leu MC, Yin Z (2018). 76:202–213. Eng Appl Artif Intell.

[4] The paper by Ananth Rao G and Kishore PVV (2017) talks about a system that can recognise Indian signs language from slow videos. Eng J Ain Shams 9(4):1929–1939.

[5] The paper by ElBadawy et al. (2017) uses 3D convolutional neural networks to recognise Arabic sign language. In: Eighth International Conference on Information Systems and Intelligent Computing (ICICIS). IEEE in Cairo.

[6] Judith Leo and Rajan, Rajesh. (2019). "A thorough examination of the system for recognising sign language."