VIVA-TECH IJRI
ANNUAL RESEARCH JOURNAL
VIVA-TECH INTERNATIONAL JOURNAL FOR RESEARCH AND INNOVATION
ANNUAL RESEARCH JOURNAL
ISSN(ONLINE): 2581-7280

REAL-TIME SIGN LANGUAGE INTERPRETER USING DEEP-LEARNING

Kriti Yadav[1], Soundarya Namal[2], Trupti Khadye[3], Madhura Ranade[4]
[1]*(EXTC, VIVA Institute of Technology, India)*
[2]*(EXTC, VIVA Institute of Technology, India)*
[3]*(EXTC, VIVA Institute of Technology, India)*
[4]*(EXTC, VIVA Institute of Technology, India)*

Abstract : *Communication is one of the basic requirements for survival in society. People with hearing or speaking impairments communicate using sign languages, and the language barrier is still a real thing. Our project aims to lessen this gap, to aid in communication, using a real-time ISL recognition system built with an LSTM algorithm. There is a lack of standard datasets for the classification of ISL characters, so we have collected a dataset using MediaPipe Holistic landmarks of face, pose, left and right hand for tracking and identifying the region of interest. The dataset consists of classes A-Z. The system collects the input via the web camera and displays the fingerspelled letter on the screen as the output. The system is trained using the LSTM algorithm and evaluated to get the best accuracy to recognize the dynamic gestures.*

Keywords – *Long Short-Term Memory (LSTM), Indian Sign Language, ISL Recognition System, MediaPipe Holistic*

I. INTRODUCTION

As said by Nelson Rolihlahla Mandela, talk to a man in a language he understands, that goes to his head. Talk to him in his own language, that goes to his heart, language is really essential to mortal commerce and has been since mortal civilization began. It's a medium humans use to communicate to express themselves and understand odds and ends of the real world. Without it, no books, no cell phones and surely not any word I'm writing would have any meaning. It's so deeply rooted in our everyday routine that we frequently take it for granted and don't realize its significance. Hardly, in the fast-changing society we live in, people with hail impairment are generally forgotten and left out. They've to struggle to bring up their ideas, voice out their opinions and express themselves to people who are different from them. Sign language, although being a medium of communication to deaf people, still has no meaning when conveyed to a non-sign language communicator. Hence, broadening the communication gap. To help this from passing, we're putting forward a sign language recognition system. It'll be an ultimate tool for people with hail disability to communicate their studies as well as a really good interpretation for non-sign language addicts to understand what the ultimate is saying. Numerous countries have their own standard and interpretation of sign gestures. For example, an ABC in Korean sign language won't mean the same thing as in Indian sign language. While this highlights the difference, it again pinpoints the complexity of sign languages. Deep literacy must be well clued with the gestures so that we can get a decent delicacy. In our proposed system, Indian Sign Language is used to produce our datasets.

The World Health Organization estimates that there are around 6.3 million people in India who have complete or partial hearing disability. Hearing loss is a spectrum, with varying types of loss and communication strategies. Some deaf people use hearing aids or cochlear implants; generally, this group chooses to lip read and use auditory cues when possible. For others, sound amplification doesn't work or is otherwise unappealing. Sign Language is the primary communication mode of communication for them. Only a minuscule percentage of hearing people understand any sign language. There has been extensive work done on American Sign Language (ASL) recognition, but not much on Indian Sign Language (ISL). ISL uses two hands for communicating and

often leads to ambiguity of features. In addition, there is a lack of ISL dataset. Our project aims to bridge this communication gap.

II. LITERATURE SURVEY

There has been work done on different sign languages around the world. Some of the earlier work that was done back in 2013 using SIFT (scale invariance Fourier transform) algorithm where scale-space feature detector was extracted through SIFT, which helped in gesture recognition with the result of 95% accuracy and 80% of the highest peak was found during testing orientation [1]. Another paper uses an image processing system to identify English alphabetic signs. First, the images are converted into Grayscale which is then converted into binary form, the accuracy was determined by matching the coordinates of the captured image to the image in the database using a comparing algorithm [2]. Recent works on recognition, for instance, ISL recognition using SVM [3]. Another SVM-related paper that is done using MediaPipe shows 99% accuracy, they have used multiple sign language datasets such as American, Indian, Italian, and Turkey for training purposes to analyze the capability of the framework [4]. Another study on ISL, using HSV Model, YIQ and YUV model, SVM, Random Forest, Hierarchical Classification [5], to check the different accuracy and to find the state of art algorithm to substitute the high-end technology like gloves or Kinect. Another popular algorithm that is highly used, when it comes to sign language recognition is CNNs. For example, an ASL recognition model, which understands finger-spelled signs was built [6]. In another paper, a system was built using CNN and PCANet for feature extraction, for the depths images that are captured from the low-cost Microsoft Kinect depth sensor [7]. Another deep learning technique approach used is Transfer Learning it is the improvement of learning in a new task through the transfer of knowledge from a related task that has already been done, here MobileNetV2 is used to get a 98.67 % accuracy [8]. In a study using Generative Adversarial Network, the proposed network architecture consists of a generator that recognizes sign language glosses by extracting spatial and temporal features from video sequences, as well as a discriminator that evaluates the quality of the generator's predictions by modeling text information at the sentence and gloss levels. A generator produces gloss predictions from video processing, while a discriminator evaluates the generator's predictions against the real gloss sequences and learns to differentiate them [9]. When dealing with sequence data, Recurrent Neural Networks is great but only to some extent. They work well for short-term dependencies, though face problems of vanishing gradient when it comes to long-term dependencies. Long short-term memory (LSTM) overcomes this problem. RNN remembers things for just small durations of time, LSTMs on the other hand, make small modifications to the information by multiplications and additions. With LSTMs, the information flows through a mechanism known as cell states. This way, LSTMs can selectively remember or forget things [10]. CNN-LSTM can be combined to construct an end-to-end network. CNN extracts the feature from input data and, LSTM layers provide sequence predictions. The result for Inception v3 CNN was 90% and Inception v3 CNN-LSTM 72% accuracy. Here, CNN gives high accuracies for isolated sign language recognition, while CNN-LSTM is a great choice for continuous word recognition [11]. Work on LSTM alone is also achieved that reaches good accuracy for some basic signs [12].

III. THEORETICAL FOUNDATION

3.1 Sign Language

All of us have different ways to communicate in order to navigate the world around us and interpret life.
Indeed, though speaking is considered the most common language mode among people, not everyone is suitable to exercise it. For someone who maintains the condition of deafness and can't hear sound, the use of audile language to change information is a no- way. A large number of the population is dissociated from the mainstream hail- dominated society and taradiddle at the threat of being marginalized, because people who are limited to using only speech can't communicate with them. A lack of availability to support the discussion between both communities also adds to the problem. Because of this, a huge challenge in the form of a communication gap between deaf, hard of hail, and hearing people arises. To bridge this gap, a non-verbal language known as sign language exists. Sign language is a non-verbal language that deaf persons simply count on to connect with their social terrain. It's grounded on visual cues through the hands, eyes, face, mouth, and body. It's a rich combination of cutlet-spelling, hand gestures, body language, facial expressions, timing, touch, and anything differently that communicates studies or ideas without the use of speech.

3.2 Long Short-Term Memory

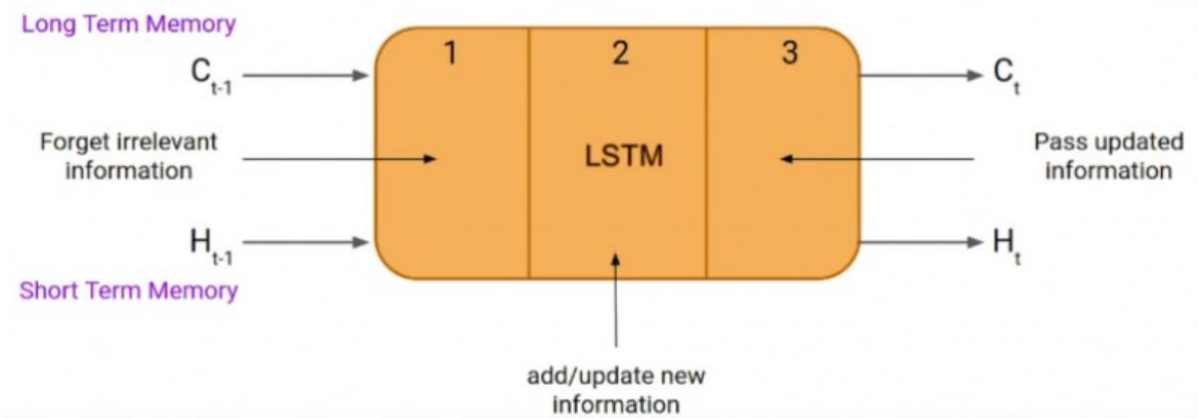

Fig 1: LSTM gates

LSTM is a special kind of recurrent neural network able of handling long term responsibilities. Long Short-Term Memory Network is an advanced RNN, a sequential network, that allows information to persevere. It's able of handling the vanishing gradient problem faced by RNN. A recurrent neural network is also called as RNN and is used for persevere memory. The RNNs remember previous info and use it to process the current input. The limitation of RNN is that they cannot remember long term responsibilities due to disappearance of the gradient. LSTM are carefully designed to keep away from long term responsibility issues.

3.2.1 LSTM Architecture

The LSTM network has three parts, as shown below and each part perform its function individually. Three parts of LSTM are called gates. The first part is Forget gate, the second part is called as the Input gate and the last one is called the Output gate.

The First gate i.e., Forget gate is used to decide whether the information from the previous timestamp should be kept or forget. Below is the equation for forget gate:

Forget Gate:

- $f_t = \sigma(x_t * U_f + H_{t-1} * W_f)$

Input gate is used to specify the importance of new information carried by input. Here is the equation of the input gate:

Input Gate:

- $i_t = \sigma(x_t * U_i + H_{t-1} * W_i)$

Output gate determines the value of next hidden state, this state have information on previous inputs.

Output Gate:

- $o_t = \sigma(x_t * U_o + H_{t-1} * W_o)$

IV. METHODOLOGY

4.1 Flow Chart

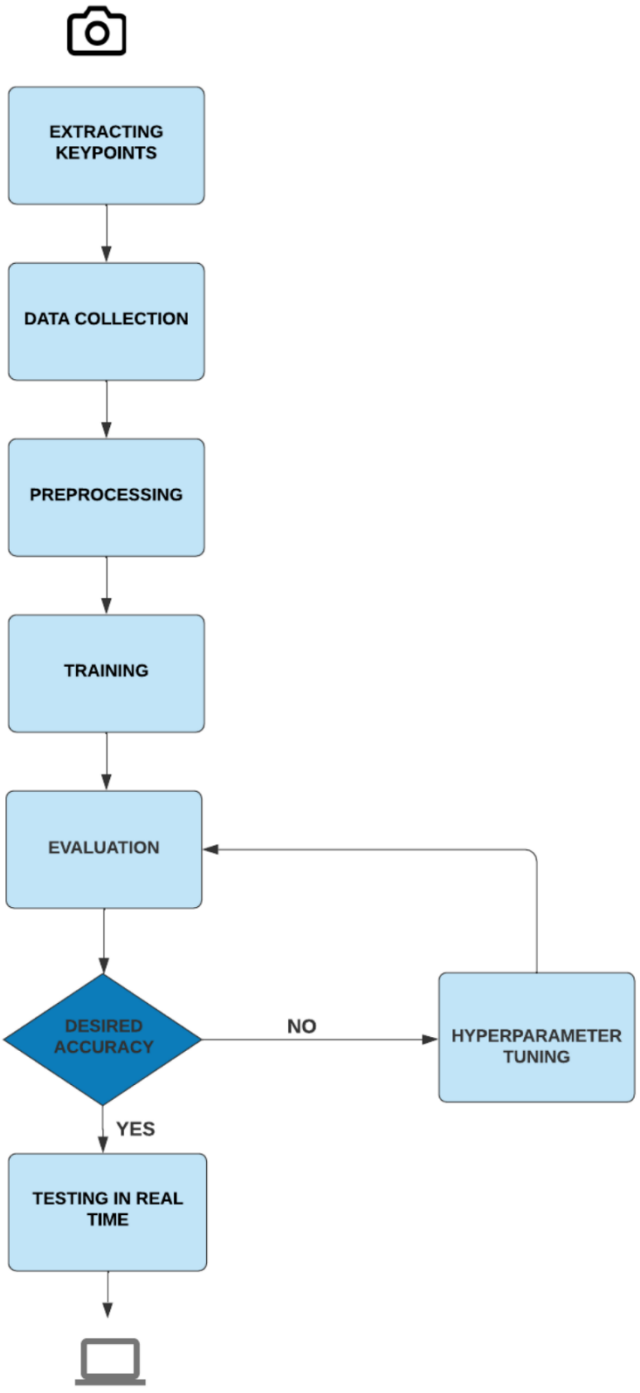

Fig 2: Flowchart

4.2 Methodology

System Overview: The overview of the proposed system is shown above. The data will be collected using a web camera and will be passed as an input of the system, the output will be a predicted sign gesture in the text on the screen. The model will be trained using the Convolution Neural Network (CNN). The system will be divided into three main phases. The first phase is the collection of data and extracting landmarks. The second phase is reprocessing i.e., making the data ready for training and, the last stage is predictions, Evaluation, and Testing in real-time. New datasets are needed to be collected due to lack of standard dataset. For data collection MediaPipe is used to track different regions of interest. The key points are extracted for the face and left and right hand. The data is collected in form of frames in order to detect the action.

Data: The dataset will be collected using a web camera having a variety of hand gestures. Collection of the data in the form of a video having frames of data. The dataset consists of classes A-Z.

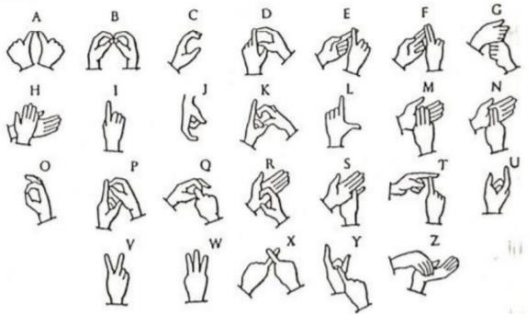

Fig 3: Alphabetic Chart in ISL

Action Detection The proposed system will be able to detect a sequence of data rather than a single frame that is used for detection. OpenCV will capture the real time hand gestures from the user and the model will detect and predict the sign and translate it into text, which will be displayed on the screen.

4.3 Tools

Language used: Python

Python is a general-purpose language, meaning it can be used to create a variety of different programs and isn't specialized for any specific problems. This versatility, along with its beginner-friendliness, has made it one of the most-used programming languages today.

Libraries used:

TensorFlow - Tensor Flow is an open-source library for ML and Artificial Intelligence. It has a comprehensive, flexible ecosystem of tools, libraries, and community resources that lets researchers push the state of the art in machine learning and developers easily build and deploy Machine learning powered applications.

OpenCV - OpenCV is a Python open-source library used for Machine Learning and computer vision in Artificial Intelligence. The library provides tools for processing and analysing the content of images, including recognizing objects in digital photos (such as faces and figures of people, text, etc.), tracking the movement of objects, converting images, applying machine learning methods, and identifying common elements in various images.

Mediapipe - MediaPipe is a framework that enables developers for building multi-modal (video, audio, any times series data) cross-platform applied ML pipelines. MediaPipe has a large collection of human body detection and tracking models which are trained on a massive and most diverse dataset of Google. As the skeleton of nodes and edges or landmarks, they track key points on different parts of the body. All coordinate points are three dimensional normalized.

Software used: Jupyter Notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modelling, data visualization, machine learning, and much more.

V. RESULTS

5.1 Action Detection

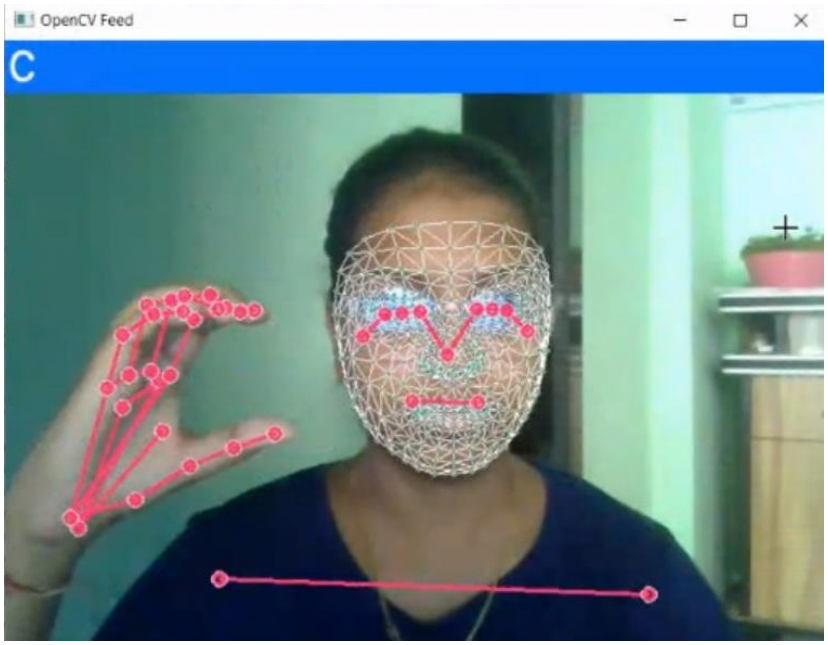

Fig 4: Detection of Letter "C"

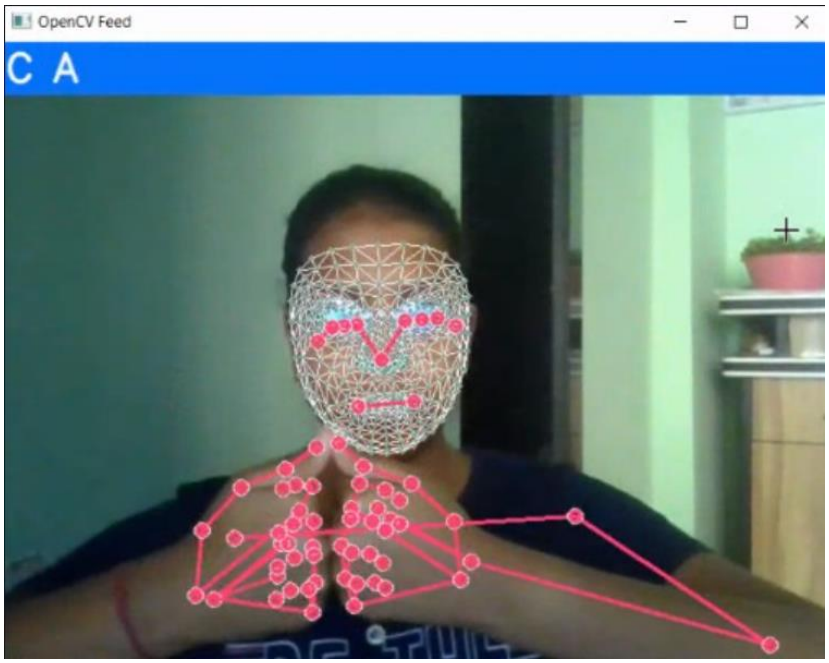

Fig 5: Detection of Letter "A"

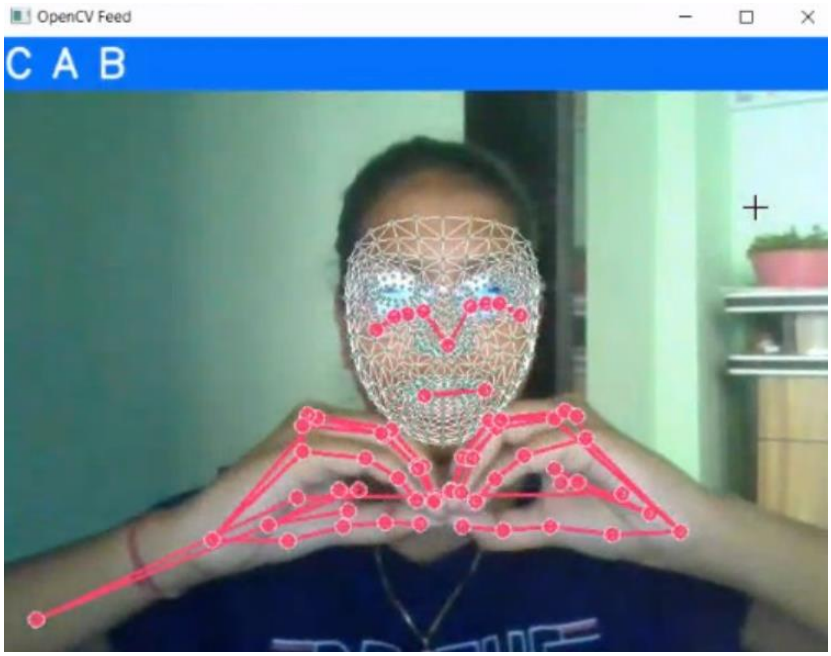

Fig 6: Detection of Letter "B"

VI. CONCLUSION

A Real-time sign language recognition system that takes input using the web camera and displays the finger spelled letter on the screen. There is a lack of for the classification of ISL characters, we have collected our dataset of classes A-Z, using MediaPipe. The system is trained using the LSTM algorithm and evaluated to get the best accuracy in order to recognize the dynamic gestures.

REFERENCES

[1] Mayuresh Keni, Shireen Meher, Aniket Mhatre, "Sign Language Recognition System", *Nation Level Students Conference on Frontiers in Engineering and Technology Application*, December 2014

[2] Sakshi Goyal, Ishita Sharma, Shanu Sharma, "Sign Language Recognition System for Deaf and Dumb People", *International Journal of Engineering Research & Technology (IJERT)* Vol. 2 Issue 4, April - 2013

[3] Prof. Radha S Shirbhate, Mr. Vedant D Shinde, Ms Sanam A Metkari, Ms Pooja U Borkar, Ms Mayuri A Khandge, "Sign Language recognition using machine learning", *International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 03*, Mar 2020

[4] Arpita Haldera, Akshit Tayadeb, "Real Time Vernacular Sign Language Recognition using MediaPipe and Machine Learning", *International Journal of Research Publication and Reviews* Vol (2) Issue (5) (2021) pp 9-17

[5] Ms Anmika Srivavtav, Mr Vikrant Malik, "Sign Language detection using Machine learning", May 2020

[6] Rasha Amer Kadhim, Muntadher Khamees, "A Real-Time American Sign Language Recognition System using Convolutional Neural Network for Real Datasets", *TEM Journal. Volume 9, Issue 3 August 2020*, pp 937-943

[7] Sanil Jain, K.V. Sameer Raja, "Indian sign character recognition", Course Project Indian Institute of Technology. Kanpur.

[8] Kin Yun Lum, Yeh Huann Goh, Yi Bin Lee, "American Sign Language Recognition Based on MobileNetV2", *Advances in Science, Technology and Engineering Systems Journal* Vol. 5, No. 6, pp 481-488 (2020)

[9] Ilias Papastratis, Kosmas Dimitropoulos, Petros Daras, "Continuous Sign Language Recognition through a Context-Aware Generative Adversarial Network", April 2021

[10] Hochreiter, S., & Schmidhuber, J., "Long short-term memory", *Neural computation, 9(8), 1735-1780,* (1997)

[11] Ahmed Adel Gomaa Elhagry, Rawan Gla Elrayes, "Egyptian Sign Language Recognition Using CNN and LSTM"

[12] Siming He, "A Sign Language Translation System Based on Deep Learning", *International Conferenceon Artificial Intelligence and Advanced Manufacturing(AIAM),* 2019