IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

HAND SIGN TO TEXT CONVERSION USING ML

[1] Prof. SONIA WAGHMARE, [2] PRAVEEN RATHOD, [3] VED SATDEVE, [4] ROHIT GAWARE,[5] RITESH KHAPARE,

[1] Lecturer in Computer Engineering Department, Marathwada Mitra Mandal's College of Engineering, Karvenagar, Pune, Maharashtra, India

[2,3,4,5] Student, Computer Engineering Department
[2,3,4,5] Marathwada Mitra Mandal's College of Engineering, Karvenagar, Pune, Maharashtra, India

Abstract: Individuals are primarily dependent on visual and auditory cues to communicate back and forth. This interaction is a two-way interaction, which means that the speaker and the listener must both take part in the conversation. However, blind and deaf people are unable to communicate in this manner. Their limitations make it difficult for them to communicate with others. Sign language is the primary method used by people who are deaf and blind. This form of communication is a set of gestures that are executed with the hands, arms, and face. The gestures are then conveyed to the other person. This form of communication is a visual and tactile form of communication that allows people who are deaf and blind to communicate with one another.

Index Terms **- Gestures, Hand, Detection, Text, Sign Language, Gesture Recognition.**

I. INTRODUCTION

Sign Language is the most natural and expressive way for individuals who are deaf to communicate. Gestures made with the hands and other body parts, including facial expressions, are included in sign language. It is primarily used by people who are deaf or dumb. Sign language recognition refers to the process of converting the user's signs and gestures into text.
Each country has its own sign language. Indians communicate using Indian Sign Language. Other sign languages, such as ASL (American Sign Language) and BSL (British Sign Language), are generally single-handed, whereas ISL uses both hands to make signs.
A deaf-dumb person's desire to speak is rarely understood by the average person. As a result, the deaf-dumb person uses gestures to communicate his or her needs. Dumb people communicate with us in their own language. In general, they communicate with others through sign language. They, however, have difficulty communicating with others who do not understand sign language.
Because new technology is being developed in our generation, we have created a machine learning computer programming model that translates sign language to text format and would reduce the communication gap between normal people and deaf people.

II. MOTIVATION

Communication is one of the main way of expressing our thoughts. The way we communicate speaks much about who we are and how we see ourselves in the society. The main goal of this project is to develop a non-verbal communication system. This project is developed for the communication between a normal person and a deaf person. The system is mainly developed for the deaf and dumb people. There are many deaf and dumb people in our country who are not able to communicate properly with the normal people.
The main objective of this project is to develop a system which will help these types of people to communicate with the normal people.

III. RELATED WORK

• In 2020 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 Algorithm Currently,In this paper four fold cross validated results for the different approaches, and the difference from the previous work done can be attributed to the fact that in our four fold i cross validation, the validation set Correspond to images of a person different from the persons in the training set.

•In 2020 Dr. Dayananda P, Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network
This research focuses on the application of a Convolutional neural network (CNN). After appropriate training, a CNN is very efficient in addressing computer vision issues and is capable of recognising the necessary characteristics with a high degree of

accuracy. The project is a straightforward instance of how CNN may be used to handle computer vision issues with exceptional precision.

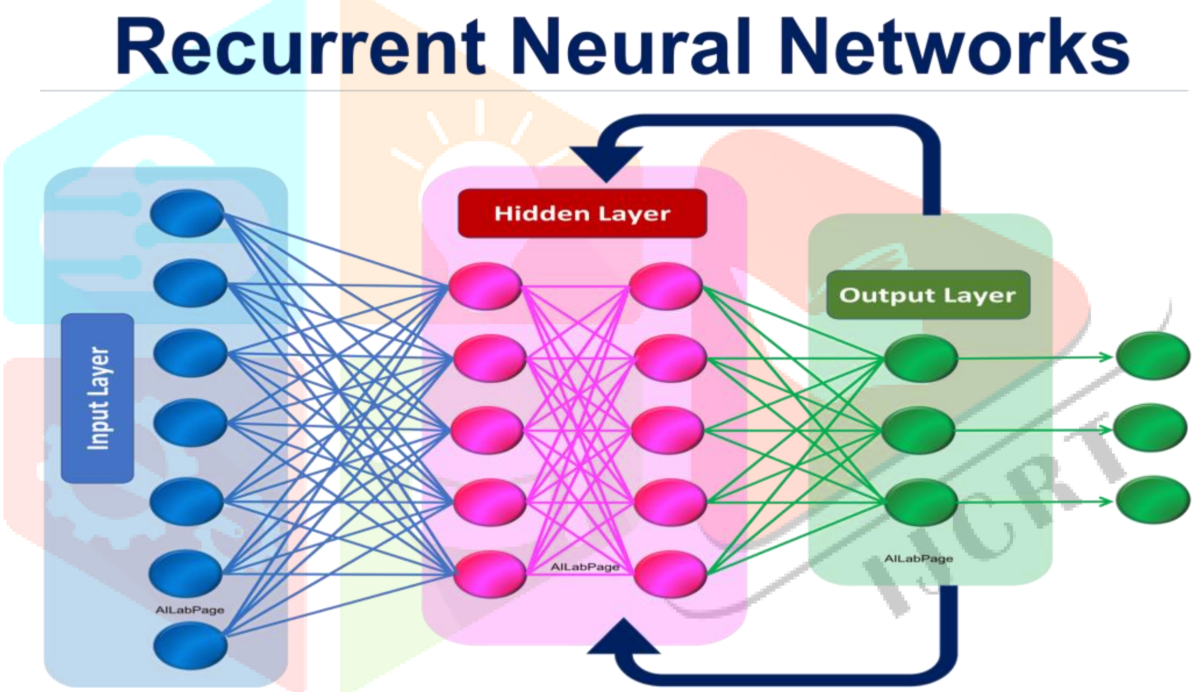
•In 2020 Hand Gesture Recognition and Voice Conversion for People Who Can't Speak Radha Abburi, R.Priyakanth, N.M.Sai Krishna. The suggested work employs a system for gathering video sequences and then extracting temporal and spatial information. "Convolutional Neural Networks" were used to train on spatial features, and "Recurrent Neural Networks" were used to train on temporal information. The datasets used are American Sign Language datasets.

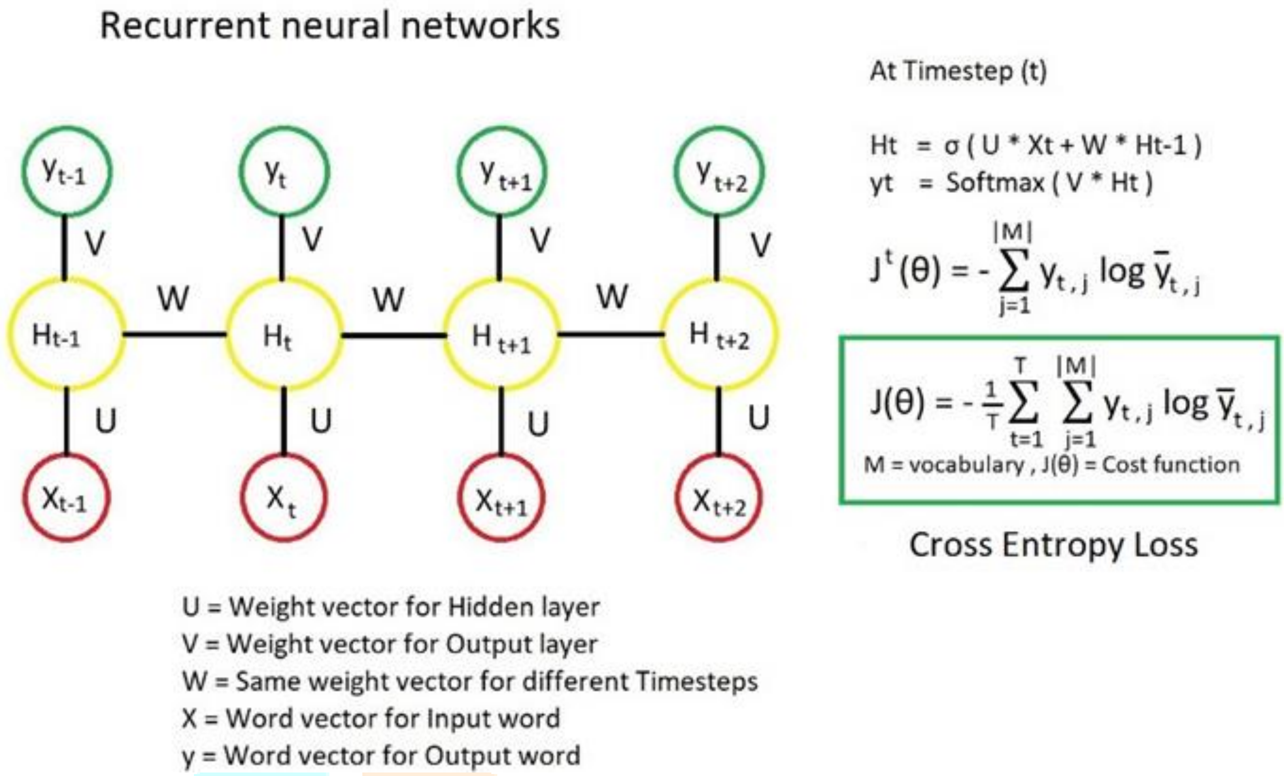
•In 2019 ,Ohnmar Win , Department of Electronic Engineering ,Mandalay Technological University, Hand Gesture to Text and Speech Conversion.This paper presents a vision-based hand gesture detection system for automated text and audio conversion utilising MATLAB software.The user can utilise the system without difficulty or complication. The application is low-cost and does not need the usage of costly technology.

IV. DATASET

No official datasets were available for Indian Sign Language and the few videos we found on internet were taken by people just describing what it looked like and not the ones that actually spoke the language. So we went to Badhirmukh Kendre, Pune a school for deaf for collecting videos of the students there making gestures for different words.
We made around a minute length video of every words taken from different students using a 30 fps camera, which roughly evaluates to around 1800 frames per words.

V. MATHEMATICAL MODEL





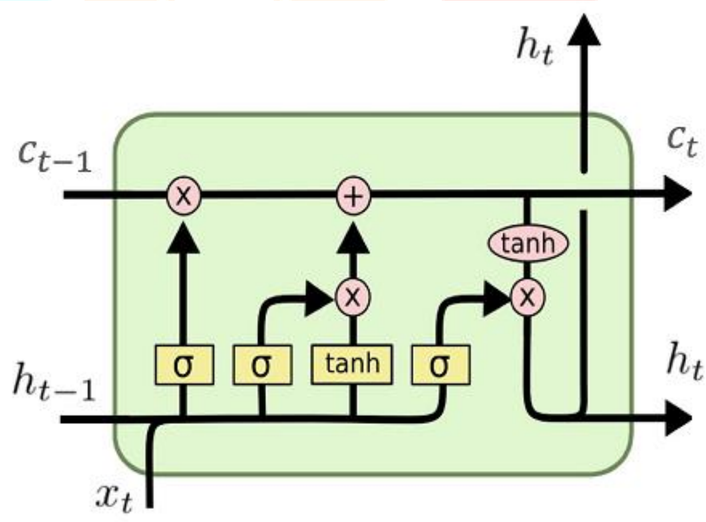


Fig. LSTM Mathematical Model

- **Worlking Of LSTM:-** The first step in our LSTM is to decide what information we're going to throw away from the cell state. This decision is made by a sigmoid layer called the "forget gate layer." It looks at h_{t-1} and x_t , and outputs a number between 0 and 1 for each number in the cell state C_{t-1}. A 1 represents "completely keep this" while a 0 represents "completely get rid of this."

Let's go back to our example of a language model trying to predict the next word based on all the previous ones. In such a problem, the cell state might include the gender of the present subject, so that the correct pronouns can be used. When we see a new subject, we want to forget the gender of the old subject.

The next step is to decide what new information we're going to store in the cell state. This has two parts. First, a sigmoid layer called the "input gate layer" decides which values we'll update. Next, a tanh layer creates a vector of new candidate values, $\tilde{C}_t$, that could be added to the state. In the next step, we'll combine these two to create an update to the state.

In the example of our language model, we'd want to add the gender of the new subject to the cell state, to replace the old one we're forgetting.

It's now time to update the old cell state, C_{t-1}, into the new cell state C_t. The previous steps already decided what to do, we just need to actually do it. We multiply the old state by f_t , forgetting the things we decided to forget earlier. Then we add $i_t * \tilde{C}_t$. This is the new candidate values, scaled by how much we decided to update each state value.

In the case of the language model, this is where we'd actually drop the information about the old subject's gender and add the new information, as we decided in the previous steps.

Finally, we need to decide what we're going to output. This output will be based on our cell state, but will be a filtered version. First, we run a sigmoid layer which decides what parts of the cell state we're going to output. Then, we put the cell state through tanh(to push the values to be between −1 and 1) and multiply it by the output of the sigmoid gate, so that we only output the parts we decided to.

$$f_t = \sigma_g \left(W_f \times x_t + U_f \times h_{t-1} + b_f\right)$$

$$i_t = \sigma_g \left(W_i \times x_t + U_i \times h_{t-1} + b_i\right)$$

$$o_t = \sigma_g \left(W_o \times x_t + U_o \times h_{t-1} + b_o\right)$$

$$c'_t = \sigma_c \left(W_c \times x_t + U_c \times h_{t-1} + b_c\right)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t$$

$$h_t = o_t \cdot \sigma_c(c_t)$$

Formula for LSTM

VI. SYSTEM ARCHITECTURE

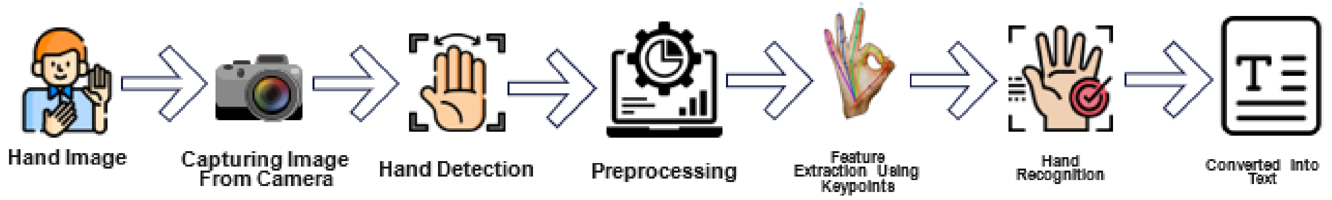


FIG 5.1 SYSTEM ARCHITECTURE

There is several steps included in conversion of hand gesture to text conversion they are

1. Hand Image:- Image acquisition is the initial stage of the vision system which is obtained by taking the snap of the image in the inbuilt camera or an external camera.
2. RGB to GRAY Conversion: Capture image is converted to RGB to gray.
3. Detection of Hand Region:- The detection of hand and the keypoints are detected by using the MediaPipe Library by using which we can detect the regions the hands and make what exactly the hand is of.
4. Preprocessing:- once the hand is detect preprocessing of the input image is done by comparing with the already trained data. In which we have used the LSTM algorithm which will keep most important data for the long term and short term data for the less important data.
5. Hand Recognition :- At this stage hand is recognized by finding the keypoints
6. Converted into Text :- At the end the hand sign is converted into text.

VII. FUTURE SCOPE

To increase user interactivity and system robustness, the application may be integrated with additional mobile devices. Using neural networks, the program's accuracy may be improved even more. In the future, numerous hand gestures may be identified and used as computer input. Hand motions indicating signals can also be turned into instructions in real time to do associated operations.
Most video games are now played on game consoles, arcade machines, or PCs, and all of them require a mix of input devices. Gesture recognition may be utilized to immerse players in the game environment in ways never previously possible.
We can utilise our machine learning application in hotels to increase the working rate of deaf individuals and provide them with a stable living. They might also have more job opportunities.

VIII. CONCLUSION

This study investigates the opportunities and limitations in hand gesture detection. Hand gesture detection is a difficult topic to solve when developing real-world apps for the deaf people. In this research, we offer an efficient approach for recognizing recorded hand movements. The backdrop was thought to be less complicated. This system does not recognize motions done with both hands. As a result, another future project may be the identification of motions done with both hands. In this project, we recognize hand gestures and successfully display text on the screen so that those who cannot understand hand gestures can communicate with deaf people.

REFERENCES

[1]. Sanil Jain and K.V.Sameer Raja, 2015 ,"Indian Sign Language Gesture recognition" .
[2]. Ohnmar Win , Department of Electronic Engineering ,Mandalay Technological University 2019,"Hand Gesture to Text and Speech Conversion".
[3]. Prof. Radha S. Shirbhate, Mr. Vedant D. Shinde , Ms. Sanam A. Metkari, Ms. Pooja U. Borkar, Ms. Mayuri A. Khandge, 2020, "Sign language Recognition Using Machine Learning Algorithm" .
[4]. Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur, Dr. Dayananda P, 2020, "Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network".
[5]. Uday Khati, Prajitesh Singh, Achyut Shankar, 2020, "Text Generation through Hand Gesture Recognition".
[6]. R.Priyakanth, N.M.Sai Krishna, Radha Abburi, 2020, "Hand Gesture Recognition and Voice Conversion for Speech Impaired".