GESTURE COMMUNICATON

Bridging ASL and text/speech

Dr.Sundara Rajulu 1#, Hemachandrika 2\*.R, Swetha.N 3#, Kiruthika.K4\*, T.V.Deepthi 5#

*Computer science Department,*

*Dhanalakshmi Srnivasan Univerrsity.*

*NH-45,Trichy-Chennai,Samayapuram,Trichy-621112*

1#drrajulu@gmail.com,2#hemachandrika12@gmail.com,3#swethan1508@gmail.com,

[4#keerthikolanji@gmail.com,5#deepthithanikonda@gmail.com](mailto:4#keerthikolanji@gmail.com,5#deepthithanikonda@gmail.com).

***Abstract*--- We offer a real-time approach that combines long short-term memory (LSTM) layers, recurrent neural networks (RNNs), and convolutional neural networks (CNNs) to recognize finger typing in American Sign Language (ASL). The technology uses sophisticated computer vision algorithms to process and classify hand movements that are captured by a camera. For hand gesture detection, MediaPipe Holistic Key points are used, and the integrated CNN-RNN-LSTM model improves accuracy and context understanding. By translating ASL into text or speech, this method successfully closes the communication gap that exists between the hearing community and deaf and hard-of-hearing people. These creative solutions are crucial for encouraging inclusive communication, as the number of people with hearing loss is expected to expand internationally.**

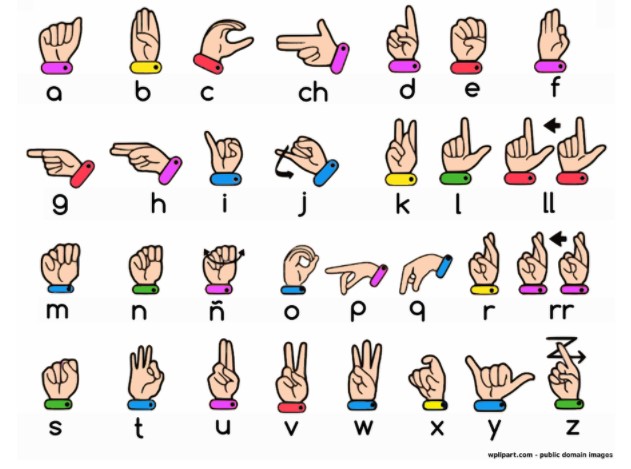
***Keywords*---**

**American Sign Language (ASL), Real-time recognition, Convolutional Neural Networks (CNNs), Recurrent Neural Networks(RNNs), Long Short-Term Memory (LSTM), Computer vision, Hand gesture recognition.**

I INTRODUCTION

Human contact requires communication in order for us to exchange ideas and feelings via voice, gestures, actions, and sights. Since traditional spoken languages are not an option for Deaf and Mute (D&M) people, sign language serves as their major means of communication. In the Deaf population in the US and abroad, American Sign Language (ASL) is extensively utilized to facilitate successful communication through hand gestures and visual signals.Our project's goal is to create a model that can identify hand movements based on fingerspelling and combine different gestures to form whole phrases. This invention, known as Gesture Comm, enables smooth communication between people utilizing various modalities by translating ASL into text and speech and vice versa.

Gesture communication facilitates understanding between disparate populations and fosters inclusivity by bridging communication gaps.

****

II LITERATURE SURVEY

Using machine learning and deep learning approaches, several studies have made substantial progress in the field of hand gesture identification for American Sign Language (ASL). Using the K-Nearest Neighbour (KNN) classifier, Dewinta Aryanie and Yaya Heryadi created a finger spelling recognition system that achieved an astounding 99.8% accuracy withfull-dimensional information for K=3. When Principal Component Analysis (PCA) was used to decrease features—principal components chosen based on Eigenvalues to account for dataset variability—their accuracy fell to 28.6%. Kshitij Bantupalli and Ying Xie extracted spatial information for sign language recognition from video streams using a Convolutional Neural Network (CNN) model called Inception. They then extracted temporal features using Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) models, and they were able to achieve 99% post-training model accuracy.Using the Marcel Static Hand Posture dataset, Tülay Karayilan and Özkan Kılıç developed neural networks for hand gesture recognition using the backpropagation algorithm. Using two classifiers—one for raw features and another for histogram features—they were able to achieve an accuracy range of 75%-85%.

In order to categorize and teach hand motions, Xinyun Jiang and Wasim Ahmad created a neural network that uses the Support Vector Machine (SVM) method. The system's capacity to differentiate five alphabets (B, D, F, L, and U)

with a success rate of roughly 99.4% was shown by their experimental results. Lastly, adding to the expanding corpus of research in this field, Galib Ibne Haidar and Hasin Ishraq Reefat from Bangladesh University of Engineering and Technology concentrated on a glove-based ASL interpretation system employing CNN and data gloves.

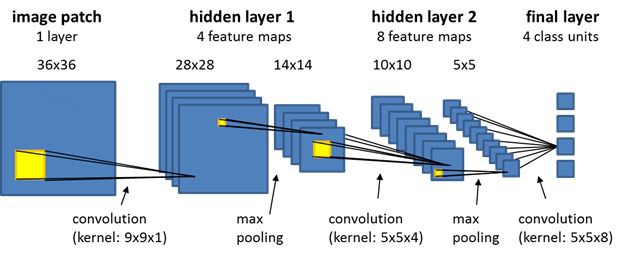
III METHODALOGY:

In this project the method we use is convolutional neural network(CNN) , Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM).

1.CONVOLUTIONAL NEURAL NETWORK(CNN): A class of neural networks called CNN is very helpful in resolving computer vision issues. Their source of inspiration was the actual visual perception that occurs in the brain's visual cortex. They employ a filter or kernel to iterate over all of the image's pixel values and perform calculations by establishing suitable weights to allow for the identification of particular features. Convolution, max pooling, flatten, dense, dropout, and a fully connected neural network layer are among the layers that CNN is composed of. When combined, these layers provide a very potent tool for feature recognition in images. Beginning layers identify low level features, which progressively start to identify higher level features that are more complicated. In contrast to conventional neural networks, CNN's layers include neurons arranged in three dimensions: width, height, and depth. Rather than being entirely coupled to every other neuron in the layer, a layer's neurons will only be connected to a tiny portion of the layer (window size) preceding it. Furthermore, as the entire image will eventually be reduced to a single vector of class scores at the end of the CNN architecture, the final output layer would have dimensions (number of classes).

CONVOLUTIONAL LAYER:

We have chosen a little window size for the convolution layer that extends to the input matrix's depth (usually 5 by 5).The layer is made up of window-sized learnable filters. In each iteration, we compute the dot product of the input values at a particular place and slid the window by the stride size, which is usually 1.As we proceed with this procedure, we will produce a 2-Dimensional activation matrix that displays the matrix's reaction at each spatial location. In other words, the network will pick up filters that turn on when it detects certain kinds of visual features, such a blotch of color or an edge with a certain orientation.

****

POOLING:

We use pooling layer to decrease the size of activation matrix and ultimately reduce the learnable parameters.

There are two types of pooling:

a. Max Pooling:

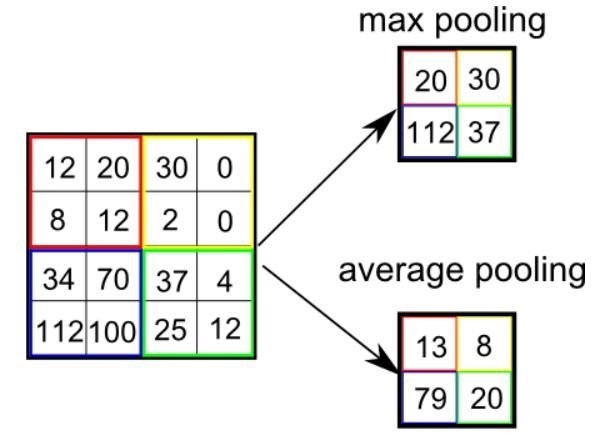
b. Average Pooling:

a .Max Pooling:

In max pooling we take a window size [for example window of size 2\*2], and only taken the maximum of 4 values. Well lid this window and continue this process, so well finally get an activation matrix half of its original Size.

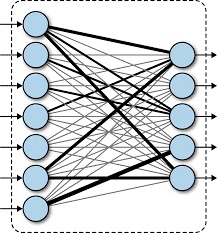
b. Average Pooling:

In average pooling we take average of all Values in a window.



**Fully Connected Layer:**

In convolution layer neurons are connected only to a local region, while in a fully connected region, well connect the all the inputs to neurons.



IV IMPLEMENTATION:

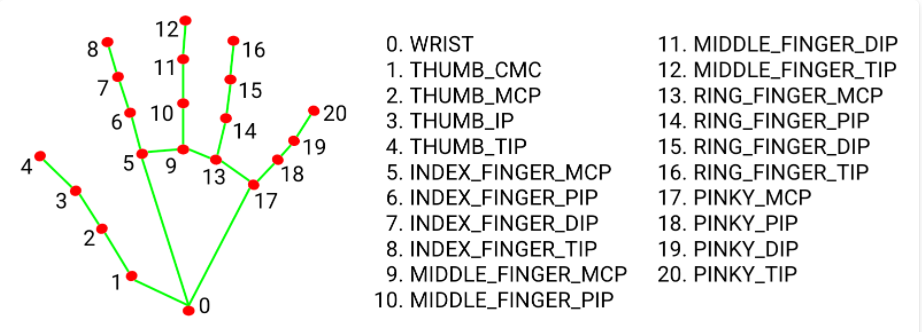
DATA ACQUISITION:

There are various approaches that can be used to get hand gesture data. One method makes use of electromechanical tools, like glove-based systems that offer accurate hand position and setup. Although these technologies are precise, they are typically costly and difficult to use. On the other hand, vision-based techniques make use of a computer webcam to track finger and hand movements. Since these techniques just need a camera to enable a natural contact between people and computers—a technology that would otherwise be necessary—they are more affordable. Vision-based techniques, however, have difficulties handling differences in skin tones, views, scales, and the speed at which the camera records a scene, in addition to variety in

hand appearance.

DATA PRE-PROCESSING:

The first step in our hand identification method is to use the MediaPipe library, which is used for image processing, to identify the hand from the webcam image. Once the hand has been identified, we use the OpenCV library to extract the region of interest (ROI), crop the picture, and convert it to grayscale. We then use a Gaussian blur filter after that. Following pre-processing, threshold and adaptive threshold techniques are used to transform the grayscale image into a binary image. For the sign letters A through Z, we gathered pictures of several signs taken from diverse viewpoints to guarantee a large dataset for testing and training.



V SYSTEM DIAGRAM:

Sign To Voice/Text

Convert into text

Train the model using RNN, CNN and LSTM Algorithm

Feature extraction

Recognize using openCV

Read the American sign language

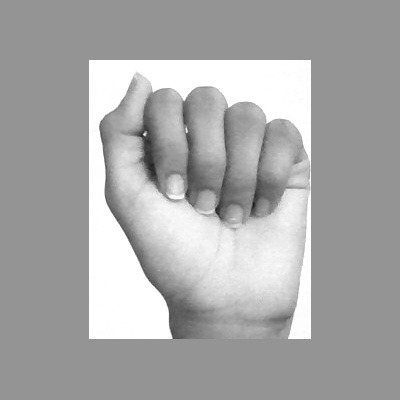
Voice To Sign

CONCLUSION:

To efficiently detect and understand American Sign Language motions from A to Z, our system combines powerful Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks. By utilizing TensorFlow's implementation power, this method guarantees the best possible performance and accuracy for real-time gesture detection. Through the use of these cutting-edge deep learning architectures, our system functions as a productive instrument to help people with impairments get beyond obstacles to communication. Future developments in this technology could lead to improvements in inclusivity and accessibility, which would benefit a variety of user populations by raising their standard of living.

RESULT:

In our project, robust real-time gesture recognition was made possible by integrating OpenCV, NumPy, Keras, MediaPipe, and TensorFlow. By applying these frameworks, we were able to read American Sign Language motions with a high degree of accuracy. This system improves the accessibility of communication for people with disabilities. Future plans call for enhancing real-time performance and investigating more extensive uses in the fields of education and healthcare.







REFERENCE:

[1] K. Dabre and S. Dholay, “Machine learning model for sign languageinterpretation using webcam images,” in International Conference on Cir-cuits, Systems, Communication and Information Technology Applications(CSCITA), 2014.

[2] D. Aryanie and Y. Heryadi, “American sign language-based ﬁnger-spelling recognition using k-nearest neighbors classiﬁer,” in 3rd In-ternational Conference on Information and Communication Technology(ICoICT), 2015.

[3] K. Bantupalli and Y. Xie, “American sign language recognition using deeplearning and computer vision,” in IEEE International Conference on BigData (Big Data), 2018.

[4] T. Karayilan and ¨Ozkan Kilic¸, “Sign language recognition,” in Interna-tional Conference on Computer Science and Engineering (UBMK), 2017.

[5] X. Jiang and W. Ahmad, “Hand gesture detection based real-time amer-ican sign language letters recognition using support vector machine,” inIEEE Intl Conf on Dependable, Autonomic and Secure Computing, IntlConf on Pervasive Intelligence and Computing, 2019.

[6] T. Rashid, Make Your Own Neural Network. CreateSpace IndependentPublishing Platform.

[7] F. Chollet, Deep Learning With Python. Manning Publications Co

[8] Sign Language to Text Conversion using Hand Gesture Recognition. Published in: 2023 8th International Conference on Communication and Electronics Systems (ICCES)

[9] Conversion of Sign Language into Text. Mahesh Kumar N B1 International Journal of Applied Engineering Research ISSN

[10] Sign Language to Text Conversion in Real Time using Transfer Learning. Shubham Thakar, Samveg Shah, Bhavya Shah and Anant V. Nimkar