iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IV **Month of publication:** April 2022

DOI: https://doi.org/10.22214/ijraset.2022.41266

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

Sign Language Recognition and Converting into Text

Shaheen Tabassum[1], Raghavendra R[2]
[1]*PG Student, Department of Master of Computer Applications, School of CS & IT, Jain Deemed-to-be-University*
[2]*Assistant Professor, School of CS & IT, Jain Deemed-to-br-University Bangalore, India*

Abstract: ***Sign language is a mode of communication that use a variety of hand movements and actions to convey a message. Deciphering these motions might be a pattern recognition challenge. People use a range of gestures and behaviours to communicate with one another. This study is a system for a human-computer interface that can identify american sign language gestures and produce textual output that reflects the meaning of the gesture. To identify and learn gestures, the proposed system would employ convolutional neural networks and long short term memory networks. This will help to break down the communication gap.***

Keywords: Data Set, Feature Extraction and Representation, Artificial Neural Networks, Convolutional Neural Networks, Tensor Flow, Keras, OpenCV.

I. INTRODUCTION

Sign language is a language for the deaf and dumb that uses simultaneous orientation and motion of hand shapes rather than acoustically conveyed different sounds.. Deaf and dumb people depend on sign language interpreters to interact. Finding competent and experienced translators for their dayto-day concerns for the entire lives, on the other hand, is a time-consuming and costly endeavor.

Sign translation is the most fundamental form of communication for persons who are deaf or hard of hearing. Those who are less fortunate endure difficulties in their everyday lives. Our idea is to produce a system that will enable interactions. Sign communication describes the use of your hands to produce shapes or motions by their connection to the head or other physical aspects, along with distinct facial features.

As a reason, a classification system must be capable of detecting various hand orientations or gestures, and also expressions and even hand position. I propose a notion for a simple but extendable system capable of distinguishing static and dynamic ASL motions, focusing on the characters a-z. It was chosen since the majority of the disabled utilize American Sign Language.

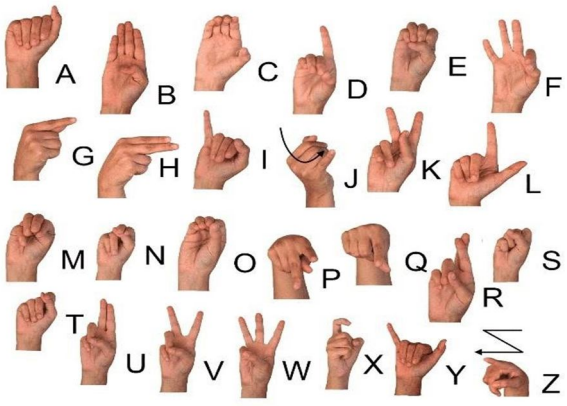


Figure 1. American Sign Language

II. EXISTING LITERATURE

Throughout my research, I came across a number of publications focusing on Translation System for Dumb and Deaf People, as well as its numerous components and methodologies.

Sakshi Goyal and Ishita Sharma (2015) create a Real - time system Identification System that collects the data, which is subsequently divided into numerous frames and characteristics such as Guassian difference in Centroids a Feature Extractor. [1]

©IJRASET: All Rights are Reserved

Iker Vazquez Lopez (2017) created Language Transcriptor, a software which interprets hand movements in photographs by applying image identification and analysis algorithms. The identification of gestures is divided into three phases: hand location, hand segmentation, and categorization. [2] Prof. Radha S. Shirbhate and Mr. Vedant D. Shinde (2020) built a system to develop a Sign Language Classification using various Computer Vision Algorithms such as SVM and KNN to produce an automated sign language sign language recognition implemented in real using multiple tools. [3]

Mohammad Elham Walizad and Mehreen Hurroo (2015) developed a Signs Language Recognition in that Convolutional Networks and machine vision are used in this system. Splitting is used to assess the whole skin tone region. Segmentation strategies are employed. Because the images generated by OpenCV are all shrunk to the same size, there is no visible difference between shots of different movements. [4]

III. PROPOSED SYSTEM

In this approach, the camera gathers images and stores them in the database under unique folders for each letter and number.

The image will be captured in RGB format, then converted to Grayscale since it just stores intensity information, making it much easier to apply a threshold to convert it to a binary image. Grayscale thresholding is then used to easily turn the pictures to binary images. For this, I'll use the Gaussian Filter because it has a median filter and is faster than others.

Thresholding is essential for reducing background noise and retaining only the hand in the image. After that, the CNN layer enters the image, matches the sequence, and changes that to texts.

There are two CNN layers in total. The very first CNN layer will categorize 26 symbols, while the second layer will classify similar-looking symbols.

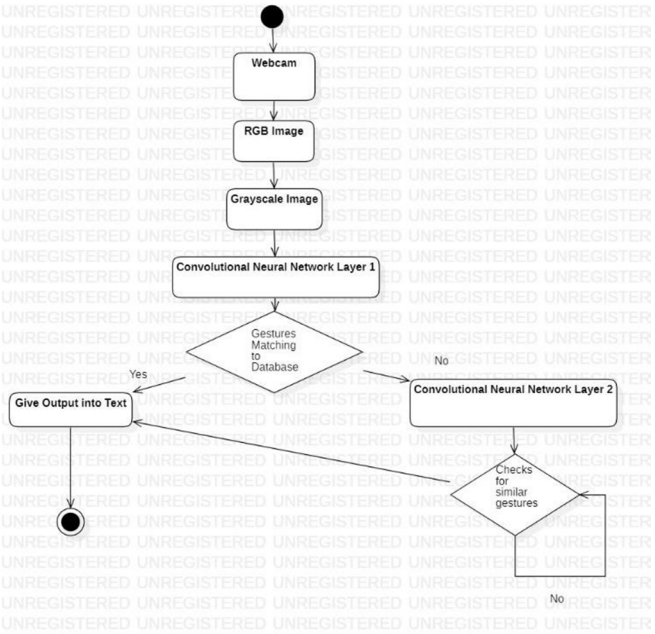


Figure 2. System Design of Sign Language Recognition System

IV. WORKING SYSTEM

The framework is based on a vision. All of the signs are represented with bare hands, which eliminates the need for any manmade gadgets for interaction.

A. Data Set Generation

In order to generate the dataset, I utilised the Open computer vision (OpenCV) package. To begin, we took around 800 photographs of each ASL symbol for training reasons and approximately 200 images of each symbol for testing purposes.

First, simply capture each frame displayed by our machine's camera. I designate an area of interest (ROI) in each frame, which is symbolized by a blue delimited square, as illustrated in the figure below. Then convert the RGB ROI from the picture sequence to monochromatic color.

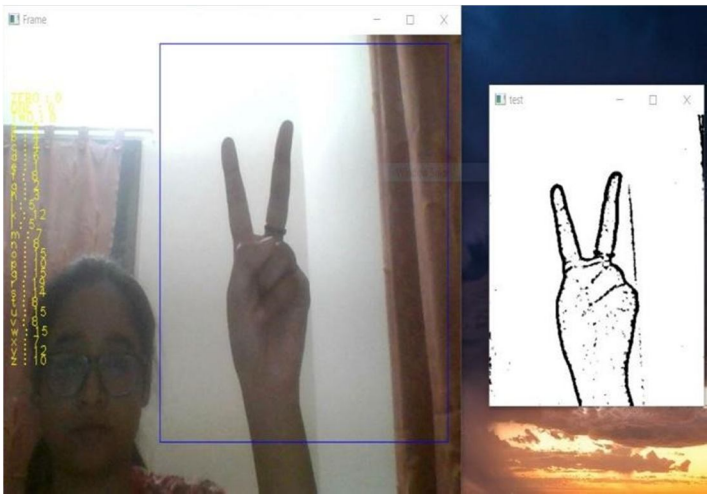


Figure 3. Capturing Raw Images



Figure 3.1. Monochrome Image

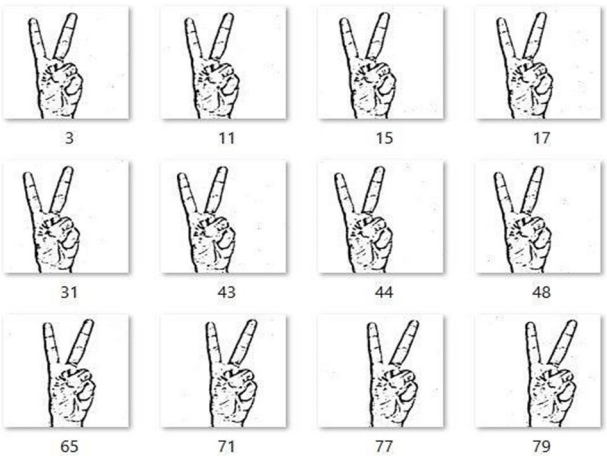


Figure 3.2. Image Post Gaussian Blur

B. Gesture Classification

Pre-processing — Colored pictures might include several characteristics, which will require a huge amount of time and resources to train the model. We can remedy this by turning the original colored image to a black and white one.

To forecast the user's final symbol, the technique employs two levels of algorithms.

1) CNN Layer 1

a) Putting the Gaussian blur filter and threshold to the OpenCV frame to obtain the transformed picture post extraction of features.

b) The transformed picture is sent to the CNN model for prediction, and if a letter is recognised for further over 50 frames, it is printed and used to build the word.

c) The blank symbol is used to represent the space between the words.

2) CNN Layer 2

I recognize several groups of letters that provide comparable consequences when detected.

And then use classifiers designed specifically for those sets to differentiate between them.

During our testing, we discovered that the following symbols were not displaying properly and were displaying additional symbols as well:

For D : R and U

1. For U : D and R
2. For I : T, D, K and I
3. For S : M and N

So, in order to handle the aforementioned instances, we created three distinct classifiers for categorizing these sets:

1. {D,R,U}
2. {T,K,D,I}
3. {S,M,N}

C. Training/Testing

To reduce superfluous noise, we transform our RGB input frames to grayscale and applying Gaussian blur. To separate our hand from the backdrop, we use adaptive threshold and scale our photos to 128 × 128. After performing all of the procedures listed above, we submit the pre-processed input photos to our model for training and testing. The prediction layer calculates the likelihood that the picture will fall into one of the classifications. As a result, the output is normalised between 0 and 1, and the total of each value in each class equals 1. We accomplished this by use the softmax function.

At start, the prediction layer's output will be considerably off from the real value. To improve it, we trained the networks with labelled data. Cross-entropy is a performance metric used in classification. It is a continuous function that is positive when the value is not the same as the labelled value and zero when the value is the same as the labelled value. As a result, we maximised the cross entropy by bringing it as near to zero as possible. To do this, we modify the weights of our neural networks at our network layer. Tensor Flow has a function for calculating cross entropy. As we discovered the cross entropy function, we optimised it using Gradient Descent; in fact, the best gradient descent optimizer is known as the Adam Optimizer.

D. Predictability

In this section, a graphical user interface (GUI) will be created. In this scenario, we'll develop a frame that will accept inputs, analyse them, and then forecast the outcomes using the model we built, which will be presented in the GUI.

V. IMPLEMENTATION

When the count of a detected letter above a certain threshold and no other letter is within a certain distance of it, we print the letter and add it to the current string (In the code we kept the value as 50 and setting the difference threshold as 20).

Otherwise, we clear the current dictionary, which contains the number of detections of the current symbol, to reduce the possibility of predicting the erroneous letter.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

When the count of detected blanks (plain backgrounds) exceeds a certain threshold and the current buffer is empty, no spaces are recognized.

In the other situation, it anticipates the end of the word by printing a space, and the current phrase is attached to the one below.

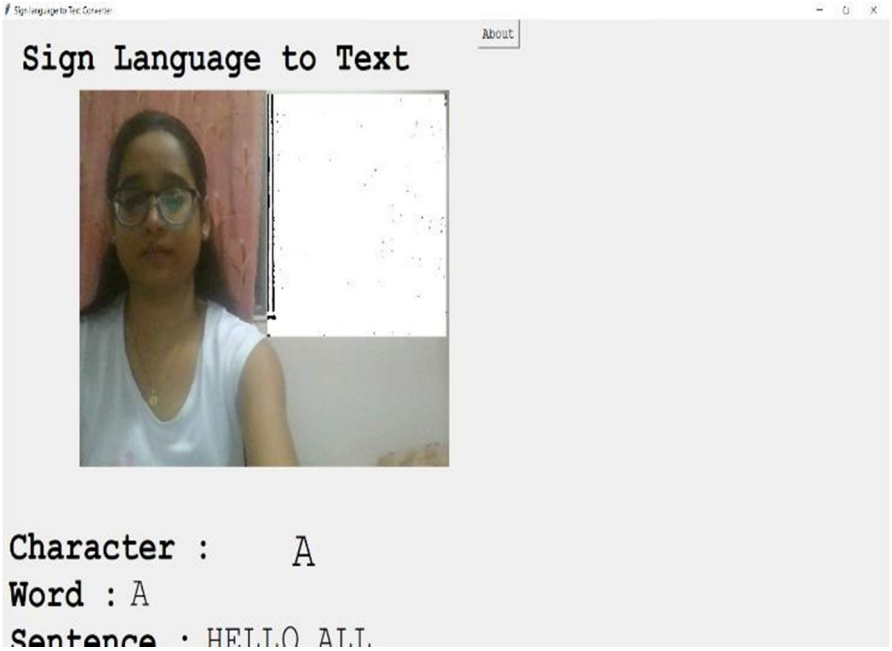


Figure 4. GUI for Translating

VI. RESULTS

I got 95.8 percent accuracy in our model utilizing only layer 1 of our technique, and 98.0 percent accuracy when layer 1 and layer 2 were merged., which is higher than the majority of existing research articles on American sign language. The majority of the research publications concentrate on the use of devices such as Kinect for hand detection.

On the other hand, while the majority of the above 21 projects make use of Kinect devices, our main goal was to design a project that could be utilized using widely available resources. A sensor like Kinect is not only not widely available, but also prohibitively costly for the majority of the audience to purchase, but our solution makes use of a standard camera on a laptop, which is a significant advantage.

The confusion matrices for the findings are shown below.

Correct \ Predicted	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	147	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	2	0	0	0	0
B	0	139	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0
C	0	0	152	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	145	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	152	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	135	0	0	0	0	0	4	0	0	0	0	0	1	0	0	2	10	0	0	0	0
G	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
H	1	0	0	0	0	0	7	143	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1
I	0	0	0	33	0	0	0	0	108	0	2	0	0	0	0	0	0	0	0	7	1	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	153	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	153	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	153	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	2	0	152	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	152	0	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	154	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	153	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	147	1	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0
S	0	0	0	0	1	0	0	0	0	0	0	0	0	1	10	0	0	0	132	0	0	0	0	8	0	0
T	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	151	0	0	0	0	0	0
U	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	0	0	115	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	151	1	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	149	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	148	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	151	0
Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Algo 1

Figure 4.1. Confusion Matrix of Layer 1

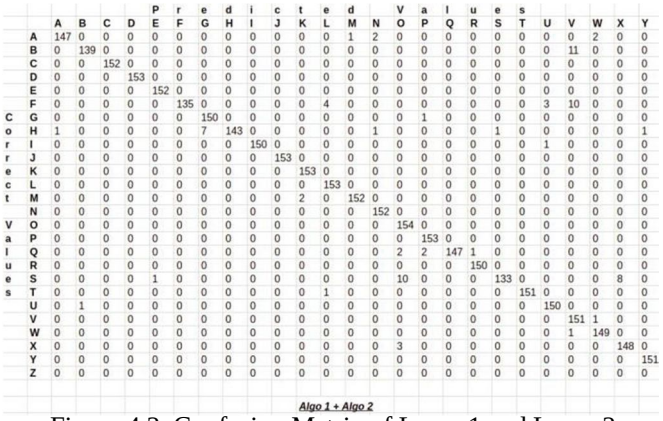


Figure 4.2. Confusion Matrix of Layer 1 and Layer 2

VII. CONCLUSION

As part of this initiative, a way for assisting dumb and deaf people in communicating more easily has been developed, and there should be no communication hurdles between us and them.

The convolution neural network's goal is to find the correct categorization. A sign language recognition system is a powerful tool for developing expert knowledge, detecting edges, and merging incorrect information from several sources.

VIII. FUTURE ENHANCEMENT

In addition to the model, by trying with various background removal methods, we hope to obtain improved accuracy even in the situation of complicated backgrounds. We are also considering upgrading the pre-processing to better anticipate gestures in low-light circumstances.

REFERENCES

[1] 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.

[2] Iker Vazquez Lopez-Hand gesture recognition for sign language transcription, Boise State University, Research and Economic Development, 2017.

[3] Prof. Radha S. Shirbhate1, Mr. Vedant D. Shinde2, Ms. Sanam A. Metkari3, Ms. Pooja U Borkar4, Ms. Mayuri A. Khandge5 "Sign language Recognition Using Machine Learning Algorithm", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 03 | Mar 2020.

[4] Mehreen Hurroo, Mohammad ElhamWalizad Sign Language Recognition System using Convolutional Neural Network and Computer Vision", International Journal of Engineering Research & Technology (IJERT), Vol. 9 Issue 12, December-2020

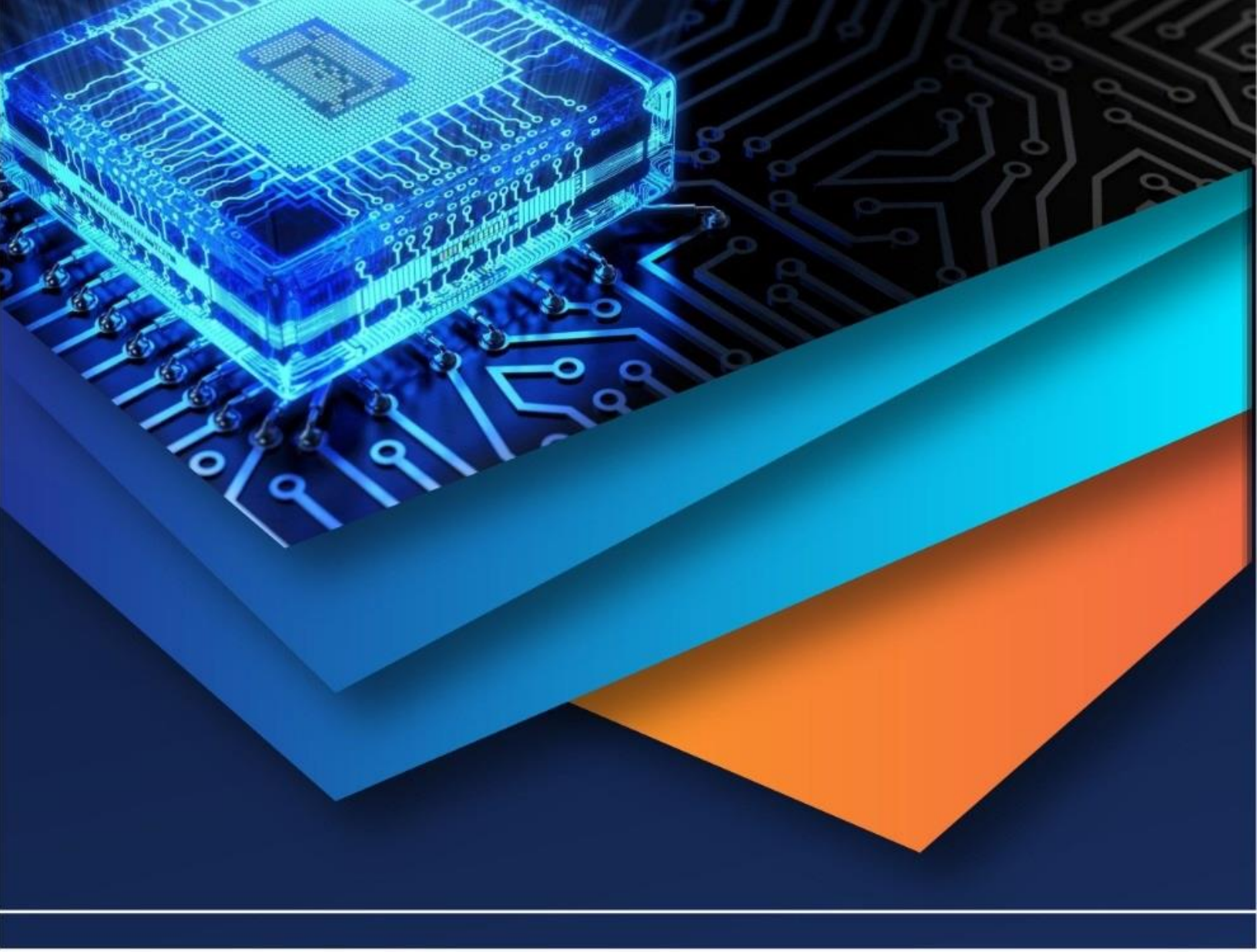
[5] Mayuresh Keni, Shireen Meher, Aniket Marathe "Sign Language Recognition System", International Journal of Engineering Research & Technology (IJERT), ICONECT' 14 Conference Proceedings

[6] H.C.M. & W.A.L.V.Kumari, & Senevirathne, W.A.P.B & Dissanayake, Maheshi, "Image Based Sign Language Recognition System For Sinhala Sign Language", Conference Paper · April 2013.

[7] A Robust Sign Language And Hand Gesture Recognition System Using Convolution Neural Networks – D Prakhya, M Sri Manjari, A Varaprasadh, NSV Krishna Reddy, D Krishna.

[8] Hemlata Dakhore , Manali Landge , Shivani Patil, Tanushree Patil , Shrutika Zyate , Ashwini Moon , Raveena Lade, "Sign Language Recognition Using Machine Learning", International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211 Volume 9, Issue 6, June -2021, Impact Factor: 7.429.

[9] Lean Karlo S. Tolentino, Ronnie O. Serfa Juan, August C. Thio-ac, Maria Abigail B. Pamahoy, Joni Rose R. Forteza, and Xavier Jet O. Garcia, "Static Sign Language Recognition Using Deep Learning", International Journal of Machine Learning and Computing, Vol. 9, No. 6, December 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)