iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: IV Month of publication: April 2024

DOI: https://doi.org/10.22214/ijraset.2024.59787

SignoSpeak: Bridging the Gap

Utsav G. Kuntalwad[1], Srushti S. Sawant[2], Mayur R. Kyatham[3], Prerna S. Shakwar[4], Dr. Varsha Shah[5]
Computer Engineering Department, Mumbai University

Abstract: "SignoSpeak: Bridging the Gap" is an innovative software research aimed at transforming communication for the hearing-impaired. It converts sign language to text in real-time, breaking down communication barriers effectively. Besides text conversion, SignoSpeak interprets various gestures, enabling users to express their thoughts vividly. By translating hand and body movements into text and audio, it enriches communication and information conveyance. This device acts as a crucial link between sign language and written language, benefiting those dependent on sign language. Additionally, its ability to recognize gestures fosters inclusive cross-cultural communication. SignoSpeak signifies a significant milestone in promoting inclusivity and meaningful interaction for the hearing-impaired. Its potential lies in revolutionizing global communication norms and understanding the unique needs of individuals with hearing impairments.

Keywords: CNN (Convolutional Neural Network), Cross-cultural communication, Hand Gestures, Hearing-impaired, Real-time conversion, SignoSpeak

I. INTRODUCTION

American Sign Language (ASL) is essential for Deaf and Dumb (D&M) individuals who cannot use spoken language for communication. ASL employs hand gestures, facial expressions, and body movements to convey meaning effectively. Unlike spoken languages, ASL is not universal and varies by region. The main objective of the research is to translate sign language to text language [3]. Efforts to bridge the communication gap between D&M and non-D&M individuals have become essential for ensuring effective interaction. Sign language is a visual language and consists of three major components:

TABLE I. Major components of visual language

Finger Spelling	World Level Sign Vocabulary	Non-manual features
Used to spell words letter by letter.	Used for the majority of communication.	Facial expressions and tongue, mouth and body position.

Depth sensors enable us to capture additional information to improve accuracy and/or processing time. Also, with recent improvement of GPU, CNNs have been employed to many computer vision problems. Therefore, we take advantage of a depth sensor and convolutional neural networks to achieve a real-time and accurate sign language recognition system [4]. In order to overcome the gap in communication caused by the difference in modes of communication, an interpreter is necessary to reduce the confusion. This research is an attempt to ease the communication between deaf and normal people. Sign language translation, a burgeoning area of research, facilitates natural communication for those with hearing impairments. A hand gesture recognition system provides deaf individuals with the means to communicate with verbal individuals independently, eliminating the need for an interpreter. Our research centres on developing a model that can accurately recognize Fingerspelling-based hand gestures, enabling the formation of complete words through the combination of each gesture. The gestures we aim to train are as given in the image below.

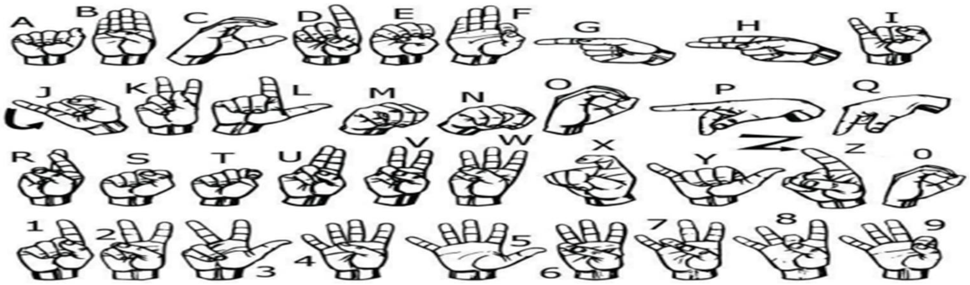


Fig. 1 ASL Hand gestures

II. THEORETICAL FRAMEWORK

A. *Data Collection*

In our research, initial attempts to find suitable datasets proved unsuccessful, as available datasets were not in the required raw image format but rather in RGB values. Consequently, we opted to create our own dataset using the OpenCV library. In order to generate the dataset, collect various images of hand signs for that here we use open computer vision (open cv) library in python to detect objects. First, we have to capture images for training and testing purpose we can capture those images through the webcam of the system[2]. The process involved capturing approximately 100 images for each of the ASL symbols A, B, and C, ensuring the dataset met our specific needs.

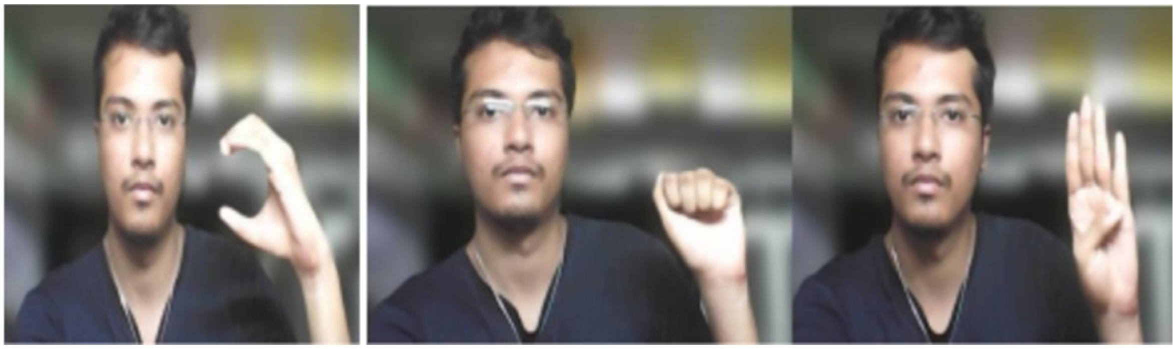


Fig. 2 Sign Recognition

B. *Data Preprocessing*

Video to Image Conversion is a pivotal step in sign language recognition, involving the extraction of frames from videos to produce still images, which are indispensable for training recognition models. Image quality is refined by denoising, contrast adjustment, and resizing, tailored to the specific needs of the dataset we captured around 800 images of each of the symbol in ASL for training purposes and around 200 images per symbol for testing purpose. First, we capture each frame shown by the webcam of our machine. In each frame we define a region of interest (ROI) which is denoted by a blue bounded square as shown in the image below. From the whole image we extracted our ROI which is RGB and convert it into grey scale Image. Finally, we apply our gaussian blur filter to our image which helps us extracting various features of our image[17]. Features extracted from the images are meticulously labeled with corresponding sign language glosses or spoken language translations, serving as target outputs for training and evaluating model accuracy. Apply Gaussian Blur filter and threshold to the frame taken with open cv to get the processed image after feature extraction[2] Accurate hand detection and landmark identification precede dataset augmentation, where techniques like rotation, translation, and flipping diversify hand poses. Extracted features are labeled with sign language glosses or translations, aiding supervised learning. These labeled datasets form the basis for training and evaluating sign language recognition models, ensuring reliable communication accessibility for users, and highlighting the importance of video-to-image conversion in system development. These meticulous processes culminate in a successful video-to-image transformation, ultimately enhancing accessibility and facilitating improved communication for users worldwide.

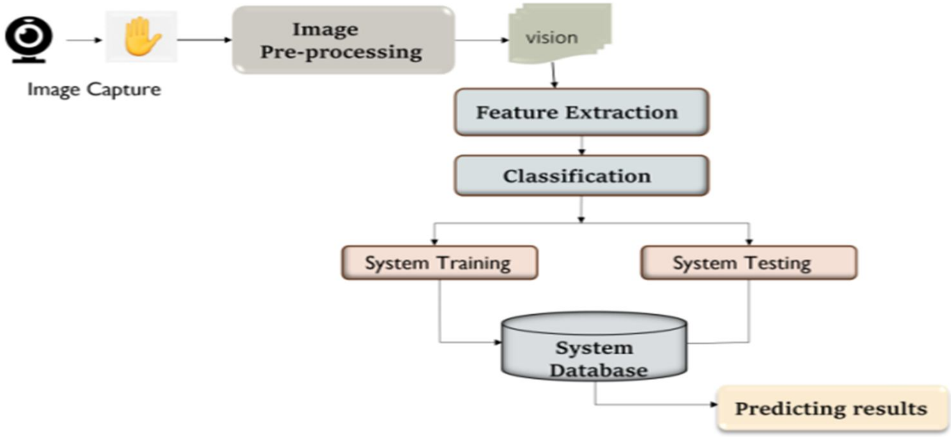


Fig. 3 Process flow

C. Train Model

Then we build a random forest classifier for hand gesture recognition task. It loads previously extracted hand landmark data, splits it into training and testing sets, trains a random forest classifier, evaluates its accuracy and saves the trained model to a file. Pickle module is used to load and save python objects as serialized binary data. Scikit-learn is used for building and evaluating the random forest classifier, as well as for splitting the dataset. Numpy is used for numerical operations. Split the data into training and testing sets using scikit-learns train_test_split function. Here 80% of the data is used for training (x_train, y_train) and 20% is used for testing (x_test, y_test) the stratify parameter ensures that the class distribution is preserved in the split. The user shows a hand gesture of any word based on the gesture language the data which is being provided in the model and also used for the testing the model [2]. Serialize and save the trained random forest model to a file called 'model.p' using pickle. This allows to use the trained model for inference without retraining it.

D. Testing

Utilizing a pretrained model, a neural network previously trained on extensive datasets, we classify hand gestures captured in real-time via a webcam using the MediaPipe library. We convert our input images (RGB) into grayscale and apply gaussian blur to remove unnecessary noise. We applied adaptive threshold to extract our hand from the background and resize our images to 128 x 128. We feed the input images after preprocessing to our model for training and testing after applying all the operations [17]. Our model actively captures frames from the webcam feed, effectively pinpointing hand landmarks through advanced algorithms integrated within the MediaPipe framework. Subsequently, based on the detected landmarks, the model makes predictions regarding the specific gestures being performed. When the user shows a gesture of any specific letter, then the model recognizes the gesture [2]. This prediction process occurs instantaneously, allowing for swift and accurate recognition of gestures as they unfold in real-time. The culmination of this process is the immediate display of the recognized gestures, providing users with timely and actionable feedback regarding their hand movements.

III. LITERATURE REVIEW

1) Paper I: Conversion of Sign Language into Text Using Machine Learning Technique [8]

This study introduces a novel approach to address communication barriers encountered by individuals who are deaf or dumb, focusing on converting hand gestures into text utilizing Convolutional Neural Networks (CNN). Given the fundamental importance of communication, particularly for those with hearing impairments, sign language serves as a vital means of expression. However, the absence of interpreters often hinders effective communication. The proposed methodology involves capturing hand gestures via camera using OpenCV and processing them through a CNN model, which includes convolution and pooling layers for feature extraction, followed by classification through fully connected layers and a Softmax output layer. The incorporation of a diverse dataset enhances the accuracy of the model. This innovative system aims to improve communication and foster inclusivity for impaired individuals by harnessing machine learning techniques and advanced image processing methodologies.

2) Paper II: Conversion of Gesture Language to Text Using OpenCV and CNN[3]

This study proposes an intuitive application to bridge communication gaps for deaf and mute individuals by translating hand gestures into text efficiently. Through the utilization of convolutional neural networks and machine learning algorithms, the system captures gestures via webcam and converts them into alphabets and words, fostering easy comprehension. A user-friendly interface enhances accessibility, featuring automatic spelling correction facilitated by Python libraries. By customizing a dataset with OpenCV to gather hand sign images and applying Gaussian blur for feature extraction, the system ensures robust gesture classification. Training the TensorFlow model involves preprocessing input images and testing for accuracy, with subsequent finger spelling implementation refining letter prediction based on gesture counts. With a final accuracy of 98.0% in real-time American Sign Language recognition, this innovative approach promises to revolutionize communication accessibility for the deaf and mute community globally.

IV. METHODOLOGY

Our application addresses system constraints by capturing webcam gestures and translating them into text using convolutional neural networks and machine learning algorithms. Through a user-friendly interface, it incorporates auto-correction functionalities via Python libraries for accurate gesture-to-text conversion. This comprehensive approach ensures seamless communication, enabling individuals unfamiliar with sign language to understand gestures effectively.

This comprehensive approach promises to enhance usability and efficiency in overcoming existing constraints. TensorFlow has an inbuilt function to calculate the cross entropy [9]. Additionally, the inclusion of word spell correction functionality enhances the precision of the model, rectifying potential errors and further refining the clarity of the conveyed message. By offering advanced features and robust functionality, our model enhances the communication experience for all users. The system is a vision-based approach. All the signs are represented with bare hands and so it eliminates the problem of using any artificial devices for interaction [17]. CNNs are inspired by the visual cortex of the human brain. The artificial neurons in a CNN will connect to a local region of the visual field, called a receptive field. This is accomplished by performing discrete convolutions on the image with filter values as trainable weights. Multiple filters are applied for each channel, and together with the activation functions of the neurons, they form feature maps. This is followed by a pooling scheme, where only the interesting information of the feature maps are pooled together [6].

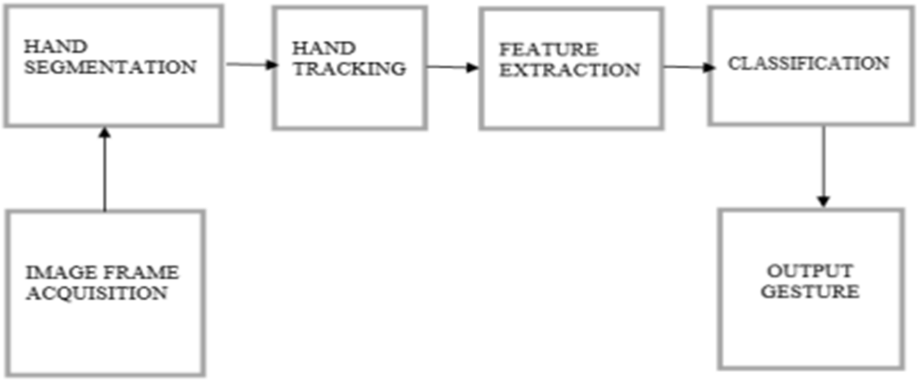


Fig. 4 Gesture detection process

Through these comprehensive features, the model contributes significantly to bridging communication gaps and facilitating seamless interaction for users with hearing impairments. PCA is a technique that allow to represent pictures as points during a low-dimensional space. If every image consists of 32x32 pixels whose values vary from zero to 255, then every image defines some points in 1024-dimensional space. If one tends to grab a sequence of pictures representing a gesture then this sequence can generate a sequence of points in space However, this set of points can sometimes lie on a low-dimensional sub-space inside the world 1024D space. The PCA algorithmic rule permits us to search out this sub space that sometimes consists of up to three dimensions. This enables us to examine the sequence of points representing the gesture [1].

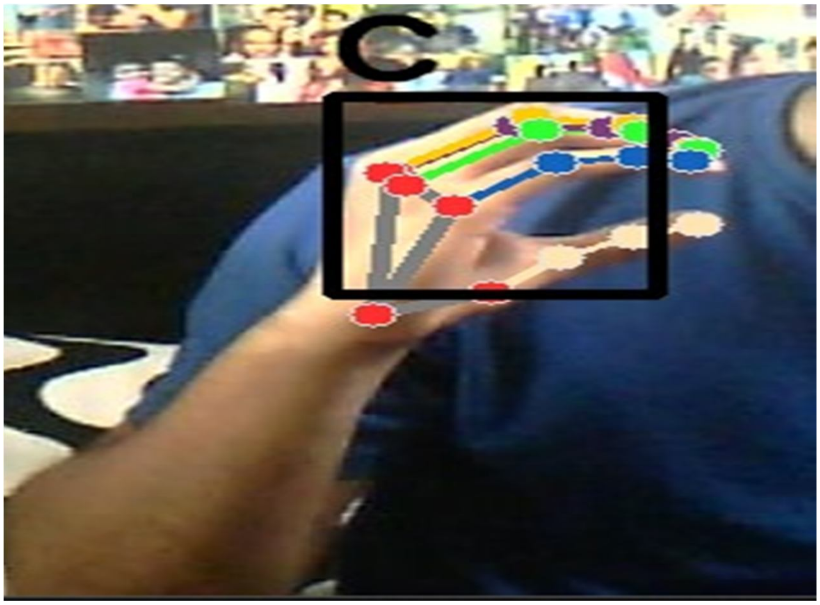
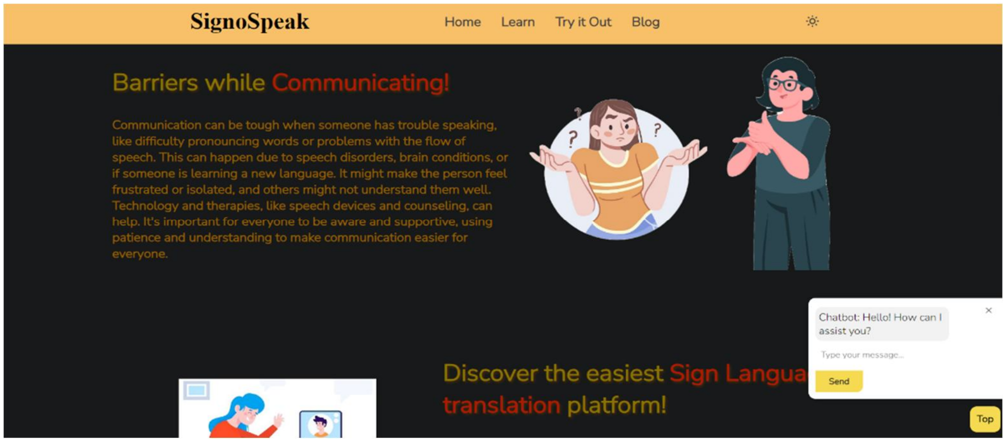


Fig. 5 Recognizing Alphabets



Barriers while Communicating!

Communication can be tough when someone has trouble speaking, like difficulty pronouncing words or problems with the flow of speech. This can happen due to speech disorders, brain conditions, or if someone is learning a new language. It might make the person feel frustrated or isolated, and others might not understand them well. Technology and therapies, like speech devices and counseling, can help. It's important for everyone to be aware and supportive, using patience and understanding to make communication easier for everyone.

Chatbot: Hello! How can I assist you?

Discover the easiest Sign Langua translation platform!

Fig. 6 SignoSpeak website

V. RESULT

In this research, a functional real time vision based American sign language recognition for Deaf and Dumb people have been developed for asl alphabets. We achieved final accuracy of 95.7% on our dataset. We are able to improve our prediction after implementing two layers of algorithms in which we verify and predict symbols which are more similar to each other. This way we are able to detect almost all the symbols provided that they are shown properly, there is no noise in the background and lighting is adequate.

VI. CONCLUSION

In our ongoing endeavour to enhance our sign language to text conversion system, we are dedicated to achieving unparalleled accuracy and precision through continuous refinement and optimization. We are expanding our dataset comprehensively to encompass a wide range of gestures, expressions, and languages, ensuring the robustness and adaptability of our system. Embracing a global perspective, we are integrating additional sign languages to promote inclusivity and accessibility across diverse linguistic and cultural backgrounds. Our focus also extends to developing a user-friendly interface that prioritizes accessibility, facilitating seamless interaction for users. As we progress, we remain committed to leveraging cutting-edge technologies to perfect our model, with the ultimate goal of empowering the global sign language community. Through our efforts, we aim to make a significant impact on individuals who rely on sign language as their primary means of communication, fostering more effective and expressive interactions worldwide. Our model provides 95.7 % accuracy for the 26 letters of the alphabet and 10 numeric digits.

VII. ACKNOWLEDGMENT

We are profoundly grateful to Dr. Varsha Shah, Principal of Rizvi College of Engineering, Mumbai for her expert guidance and continuous encouragement throughout to see that this research rights its target.

We would like to express deepest appreciation towards Dr. Anupam Choudhary, Head of Department and Prof. Shiburaj Pappu, Dean of Computer department whose invaluable guidance supported us in this research.

At last, we would express our sincere heartfelt gratitude to all the staff members of computer engineering department who helped us directly or indirectly during this course of work.

REFERENCES

[1] T. Yang, Y. Xu, and “A., Hidden Markov Model for Gesture Recognition”, CMURI-TR-94 10, Robotics Institute, Carnegie Mellon Univ., Pittsburgh, PA, May 1994.

[2] Pujan Ziaie, Thomas M uller, Mary Ellen Foster, and Alois Knoll “A Na ̈ive Bayes Munich, Dept. of Informatics VI, Robotics and Embedded Systems, Boltzmannstr. 3, DE-85748 Garching, Germany.

[3] Mohammed Waleed Kalous, Machine recognition of Auslan signs using PowerGloves: Towards large-lexicon recognition of sign language.

[4] Byeongkeun Kang, Subarna Tripathi, Truong Q. Nguyen” Real-time sign language fingerspelling recognition using convolutional neural networks from depth map” 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)

[5] ijraset.com/best-journal/sign-language-interpreter

[6] Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. (2015) Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8925. Springer, Cham

[7] Zaki, M.M., Shaheen, S.I.: Sign language recognition using a combination of new vision-based features. Pattern Recognition Letters 32(4), 572–577 (2011)

[8] ijraset.com/best-journal/conversion-of-sign-language-video-to-text-and-speech

[9] N. Mukai, N. Harada and Y. Chang, "Japanese Fingerspelling Recognition Based on Classification Tree and Machine Learning," 2017 Nicograph International (NicoInt), Kyoto, Japan, 2017, pp. 19-24. doi:10.1109/NICOInt.2017.9

[10] Number System Recognition (https://github.com/chasinginfinity/number-sign-recognition)

[11] Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D., Bengio, Y.: Theano: a CPU and GPU math expression compiler. In: Proceedings of the Python for Scientific Computing Conference (SciPy), June 2010, oral Presentation

[12] “Sign language recognition.” In visual Analysis of Humans, pp. 539- 562. Springer London,2011. Cooper, Helen, Brian Holt, and Richard Bowden.

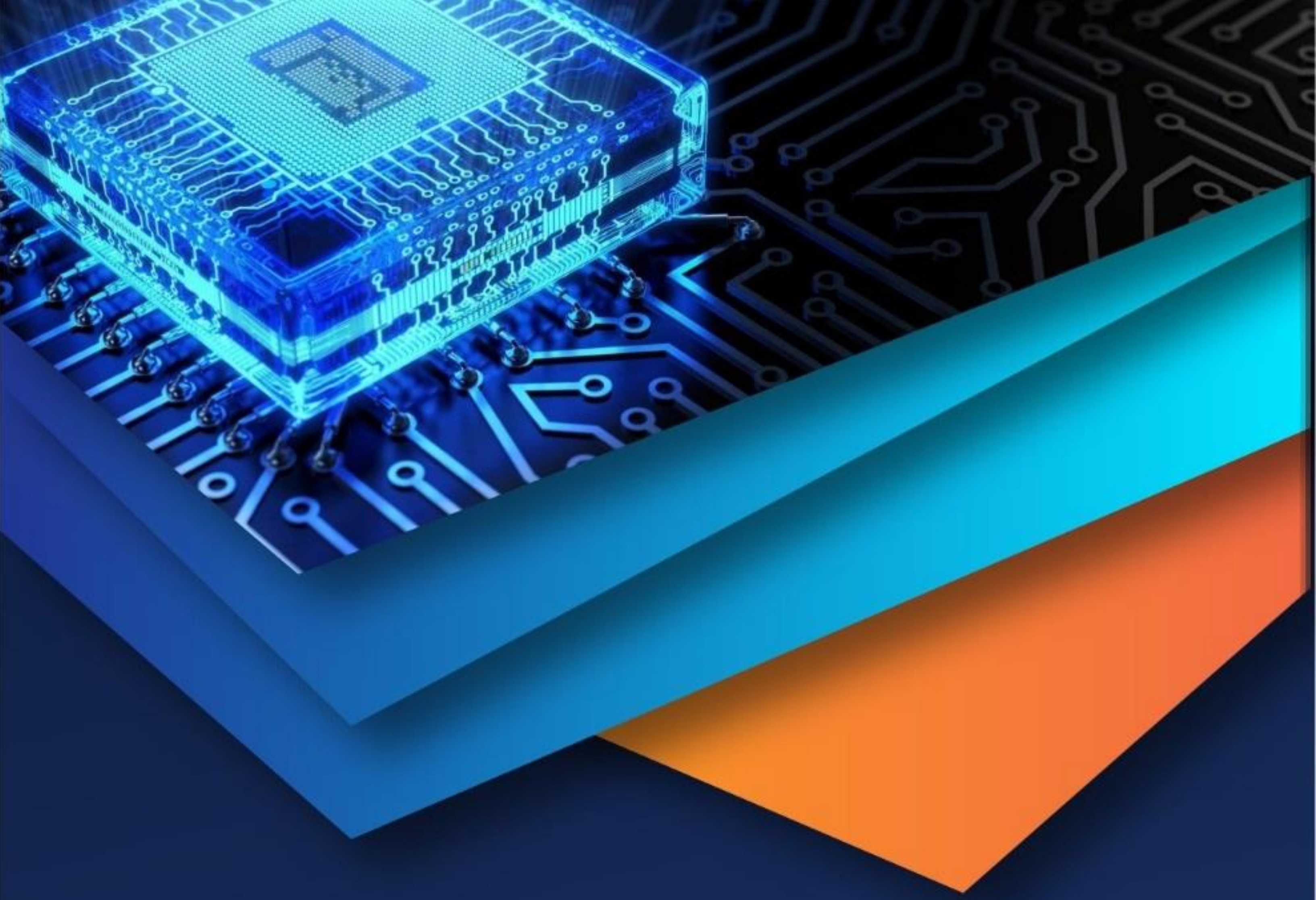
[13] “Handshape recognition for Argentinian sign language using problem”. Journal of Computer Science and Technology 16(2016). Ronchetti, Franco, Facundo Quiroga, Cesar Armando Estrebou and Laura Cristina Lanzarini.

[14] T. Starner, J. Weaver, and A. Pentland. Real-time american sign language recognition using desk and wearable computer based video. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 20(12):1371–1375, Dec 1998

[15] S. Liwicki and M. Everingham. Automatic recognition of fingerspelled words in british sign language. In Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on, pages 50–57, June 2009.

[16] J. Isaacs and S. Foo. Hand pose estimation for american sign language recognition. In System Theory, 2004. Proceedings of the Thirty-Sixth Southeastern Symposium on, pages 132–136, 2004

[17] https://www.ijert.org/a-review-paper-on-sign-language-recognition-for-the-deaf-and-dumb



doi crossref
10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

SJIFactor
TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY