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.2024.60038

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 12 Issue IV Apr 2024- Available at www.ijraset.com

Sign Language Recognition using CNN

Shreya Prabhupatkar[1], Gaurav Singh[2], Pranav Patil[3], Prashant Singh[4], Prof. Nishikant Khaire[5]
Department of Information Technology, Datta Meghe College of Engineering, Airoli, Navi Mumbai

Abstract: ***The sign language Recognition System is a technology that bridges the gap between the deaf people and the hearing world. Sign language is a vital mode of expression for deaf people, yet effective communication remains a challenge. This project aims to develop a robust Sign Language Recognition System capable of accurately translating sign language gestures into text. The system utilizes a deep learning approach, specifically convolutional neural networks (CNNs) to analyze video input of sign language and then generate the corresponding output. The project involves data collection, pre-processing, segmentation, feature extraction, model training, and evaluation. The proposed SLR system has the potential to enhance. Communication and accessibility for deaf individuals, promoting inclusivity and improving their quality of life. The project represents a comprehensive exploration of both technical and ethical aspects in the realm of computer vision and deep learning applications.***
Keywords: Sign language recognition, Convolutional Neural Network (CNN), Deep learning, Computer vision, Real-time recognition, Gesture recognition, Machine learning.

I. INTRODUCTION

Communication is a fundamental human right. It helps two individuals or groups of individuals to share their thoughts and information, whether it is verbal or non-verbal communication. However, there are still some individuals with hearing impairments and mute people who have difficulty expressing themselves. They converse with each other using sign language, but it's very uncommon that the other person they wish to talk knows the sign language. The world has advanced greatly in recent years. Sign language recognition has emerged as a beacon for hearing-impaired people, and it bridges the gap between the hearing-impaired community and the rest of the people so that they can communicate with each other without having any trouble. In this project, we have combined computer vision and information technology to detect objects and gestures using the MediaPipe library. Our system processes the gesture once an object has been detected and turns it into text. This makes it easy for the other user who is having trouble understanding the sign language.

The Sign Language Recognition System (SLRS) project represents a pioneering effort to bridge the communication gap between individuals who use sign language as their primary means of expression and those who rely on spoken or written language for communication. Sign language is a crucial mode of interaction for the Deaf and hard-of-hearing community, serving as a foundation for conveying thoughts, emotions, and information. However, the unique visual and gestural nature of sign language poses a challenge for those who are not proficient in it.

The primary objective of the SLRS (Sign language recognition system) project is to develop a comprehensive solution that can accurately recognize and interpret sign language gestures in real-time. This recognition is subsequently translated into natural language text or spoken language, enabling seamless communication between sign language users and individuals who may not understand sign language.

This project is motivated by the pursuit of inclusivity and accessibility for all members of society. It seeks to address the longstanding communication barriers faced by the Deaf and hard-of-hearing community, which have implications in diverse domains such as education, healthcare, customer service, and everyday interpersonal interactions.

Advancements in technology, particularly in the fields of machine learning, computer vision, and natural language processing, have created a fertile ground for the development of a robust Sign Language Recognition System. The system's success would signify a significant leap towards realizing equal access to information and services for all, irrespective of their preferred mode of communication.

II. LITERATURE SURVEY

In the paper titled "Marathi Sign Language" Swaraj Dahibavkar [2] propounded that the input is processed to detect the object. Segmentation is performed on the object to differentiate the hand from the background. Edge-based segmentation is used to detect the edge, and then the edges are connected to form the boundaries. To detect the edges, they have used basic edge detection techniques such as the Sobel operator, canny operator, Robert operator, etc. and to train the machine, they have used the CNN algorithm, which helps assign the important object in the image to differentiate one from another.

©IJRASET: All Rights are Reserved

In the paper titled "Sign Language Recognition for Deaf and dumb people" Y.M. Pathan and S.R. Waghmare [4] proposed the four major steps, which are Skin fileting helps extract the skin-coloured pixels from the non-skin-coloured pixels. Then the input image is converted into the HSV image because RGB images are very sensitive to changes in illumination conditions. The HSV colour separates the three components: Hue, which means the set of pure colours within colour space; saturation describes the purity of a colour image; and value gives the lightness and darkness of a colour. Then the HSV image is filtered and smoothed. Along with the hand, the surrounding area may also contain the skin-coloured object. To eliminate that, they used the biggest binary linked object (BLOB), which considers only the region comprising the biggest linked skin-coloured pixels. In the second step, they describe hand cropping, in which the unnecessary part is clipped off using the hand cropping technique because, for gesture detection, only the hand-filled wrist part is required. In the third step, they extract the feature from the clipped hand, and in the last step, they recognize the sign with the help of the extracted features.

In the paper titled "Indian Sign Language Recognition using Convolutional Neural Network" Rachana Patil and Vivek Patil [6] proposed the segmentation of the input image to separate the hand from the background, and then they extracted the feature based on contour, geometrical feature (position, angle, distance, etc.), colour feature, etc. After the feature extraction, they extract the best features from it to reduce the computational work, and then using the best features, they predict the gestures.

III. PROPOSED METHODOLOGY

1) Image Acquisition: In this process, an image is acquired with the help of a camera or webcam. Nowadays webcam comes inbuilt into the laptops which help to capture images and record videos. Gestures can be captured using this camera and it is given as input to the next phase.

2) Skin filtering: In this phase, skin is filtered from the input image by filtering skin-coloured from non-skin coloured. This method is very useful for the detection of hand and face etc.

3) Hand Cropping: In this phase, the input image is cropped around the hand portion to only include the hand area till the wrist, and unnecessary parts are clipped off, which helps to locate the fingers to recognize the gesture.

4) Segmentation: The method of separating objects or signs from the context of a captured image is known as segmentation. Context subtracting, skin-color detection, and edge detection are all used in the segmentation process. The motion and location of the hand must be detected and segmented in order to recognize gestures.

5) Feature Extraction: To extract the features from the input image, we are making use of the MediaPipe library: MediaPipe is an open-source framework for building pipelines to perform computer vision inference over arbitrary sensory data, such as video or audio. Using MediaPipe, a perception pipeline can be built as a graph modular component, including, for instance, models and media processing functions. Sensory data, which is a frame-by-frame image from the videos, enters into the graph, which helps to draw the landmark on the hand. Graphs of operations are used, such as tensorflow, to define a neural network. In MediaPipe (Convolutional Neural Network) CNN is integrated for feature extraction.

6) Sign Recognition: In CNN the input goes through several layers and in the end last layer which is the output layer, after performing several calculations, it generates the result and it is shown in the form of text which is the translation of the gesture.

7) Model Evaluation: Split the dataset into training and testing sets. Make adjustments to the model architecture or hyperparameter if necessary.

8) Model Deployment: Integrate the trained model into a desired application or system. Implement any necessary post-processing steps, such as filtering out predictions with low confidence scores. Continuously monitor and update the model as needed to maintain its accuracy.

IV. WORKING OF CNN

A. Convolutional Neural Network (CNN):

CNN is a type of artificial neural network that uses a Deep learning method, as the name suggests it goes deep down several networks to extract the important and hidden features from the image. It contains three layers: an input layer, an output layer, and various hidden layers:

1) Input Layers: The layer in which input is given to our model. The number of neurons in this layer is equal to the total number of features in our data.

2) Hidden Layer: In this layer the input from the Input layer is fed into the hidden layer. There are many hidden layers between the input and the output layer. Each hidden layer can have different numbers of neurons which are generally greater than the number of features.
3) The output from each layer is computed by matrix multiplication of the output of the previous layer with learnable weights of that layer and then by the addition of learnable biases followed by activation function which makes the network nonlinear.
4) Output Layer: The output from the hidden layer is then fed into a logistic function like the sigmoidal or softmax layer which converts the output of each class into the probability score of each class.

The Human Brain performs a large-scale convolution to process the visual signal received by the eyes, based on the observation. We have also built a similar system that will take visual input through the camera and then perform certain actions on the input image to give the result as an output. To perform actions on the input image we are using CNN. In CNN two major operations are performed which are convolution (wT*X) and pooling (max ()) and these blocks are wired in a highly complex fashion to mimic the human brain. The neural network is constructed in layers, where the increase in the number of layers increases the network complexity and is observed to improve the system accuracy. The CNN architecture consists of three operational blocks which are connected as a complex architecture:

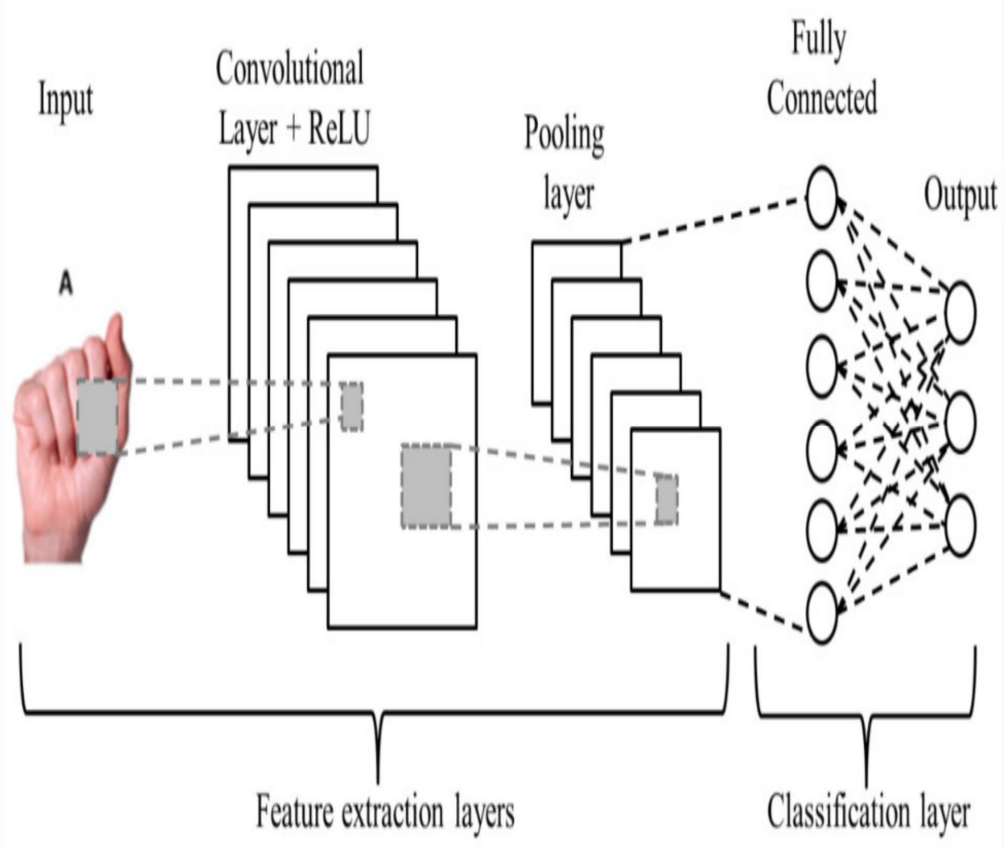


Fig. 1 Convolutional Neural Network Architecture

B. The functional block of Convolutional Neural Network (CNN):

1) Convolution Layer: The convolutional layer is the fundamental building block of a CNN and is where the majority of computations occur. This layer uses a filter or kernel (a small matrix of weights) to move across the receptive field of an input image to detect the presence of specific features.
2) Pooling Layer: The pooling layer aims to reduce the dimensionality of the input data while retaining critical information, thus improving the network's overall efficiency.
3) Fully Connected Layer: The fully connected layer plays a critical role in the final stages of a CNN, where it is responsible for classifying images based on the features extracted in the previous layers. The term fully connected means that each neuron in one layer is connected to each neuron in the subsequent layer.

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

C. System Diagram:

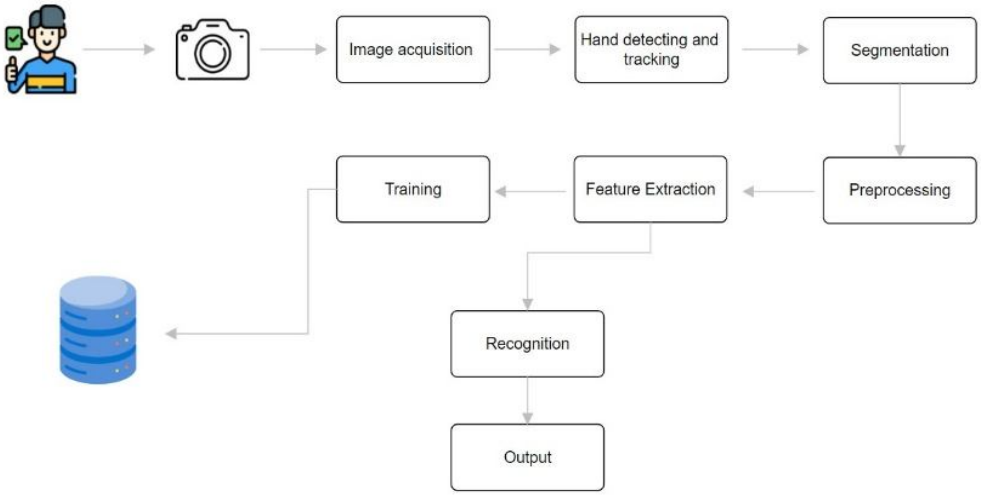


Fig. 2 Data Flow

V. EXPERIMENTATION

To create a real-time Sign Language Recognition System we went through the following steps:

1) Project Scope and Objective: The objective of our project is to capture sign language in real-time and interpret it into textual form.

2) Data Collection and Preparation: Generating a diverse dataset of sign language gestures, ensuring representation of various signs and hand movements, and pre-processing the data to remove noise and prepare it for training and testing.

3) Image Capture Setup: Set up the camera-enabled device and configure OpenCV to capture high-quality images from signers. Ensuring the camera is positioned correctly for optimal hand tracking.

4) Landmark Detection Implementation: Implement hand landmark detection using the cvzone library. Verifying that the library accurately tracks the landmarks on the signer's hands and captures real-time movement.

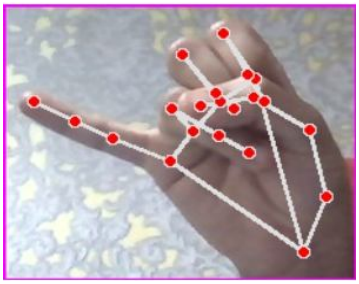


Fig. 3.1 Landmark Detection I

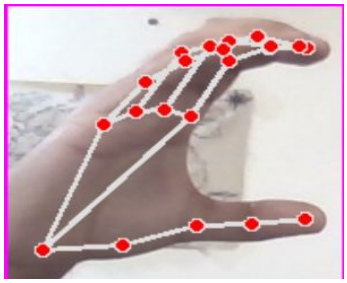


Fig. 3.2 Landmark Detection II

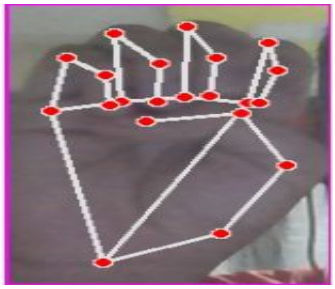


Fig. 3.3 Landmark Detection III



Fig. 3.4 Landmark Detection IV

©IJRASET: All Rights are Reserved

5) Feature Extraction: Developed algorithms to extract relevant features from the detected landmarks. Features include hand position, finger configurations, and motion trajectories
6) Model Training and Testing: Trained the selected model on the pre-processed dataset. Evaluated the model's performance using appropriate metrics and conducted rigorous testing to ensure accuracy and generalization
7) Real-Time Gesture Recognition: Integrated the trained model into the real-time video processing pipeline. Also ensured that the system can interpret sign language gestures on-the-fly and generate textual output.

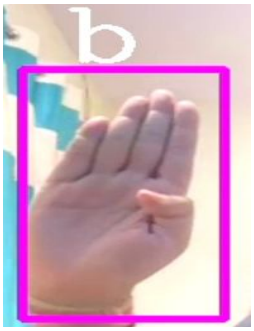


Fig. 4.1 Gesture Recognition I



Fig. 4.2 Gesture Recognition II



Fig. 4.3 Gesture Recognition III



Fig. 4.4 Gesture Recognition IV

VI. CONCLUSIONS

In conclusion, the development of a sign language detection system represents inclusivity and accessibility for hearing-impaired people. Through the use of machine learning algorithms, particularly deep learning models like convolutional neural networks (CNNs), it is possible to accurately translate sign language gestures into text or speech, thereby enabling more effective communication between deaf individuals and the hearing world.

The successful implementation of a Sign Language Recognition system requires careful consideration of factors such as data collection, segmentation, pre-processing, feature extraction, model training, and evaluation. Additionally, the system should have lower latency and a user-friendly interface.

Overall, the development of a Sign Language Recognition system has the potential to significantly improve the quality of life for deaf individuals by providing them with a more effective means of communication. The system also takes feedback from the user to understand the user requirements and improve itself based on the feedback to improve accuracy and usability. The sign Language Recognition System represents a crucial step towards breaking down communication barriers and promoting greater understanding and inclusion for all.

VII. FUTURE SCOPE

The successful development of the Sign Language Detection system opens up a wide array of promising avenues for future work and enhancements. First and foremost, further expanding the system to recognize and interpret sign language from different regions and languages is a significant opportunity. Sign languages exhibit notable variations across the globe, and accommodating these differences will make the system more inclusive and globally accessible.

Additionally, the integration of real-time translation features into mobile applications and wearable devices could profoundly impact the daily lives of individuals who use sign language.

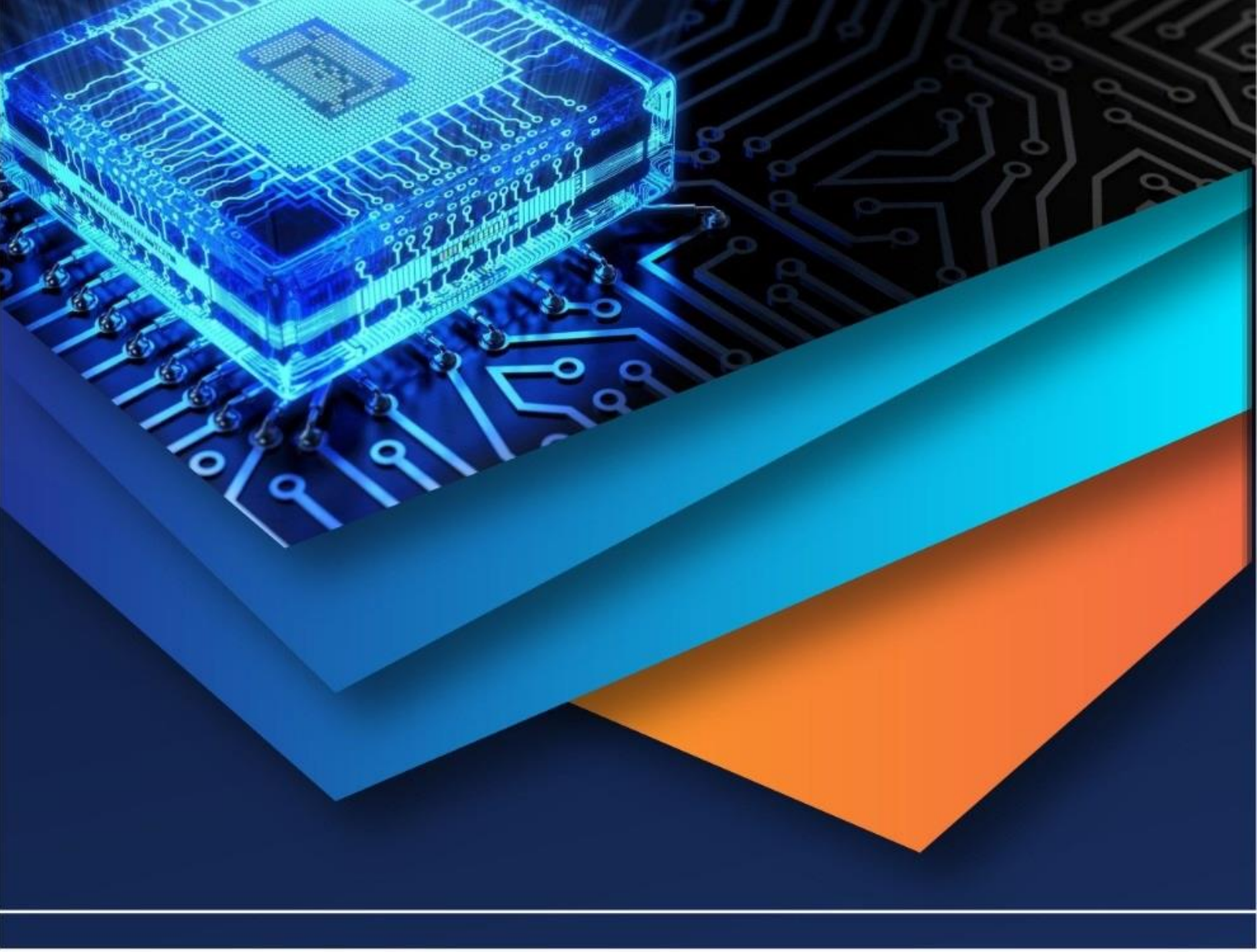
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Imagine a smartphone app that not only detects and interprets sign language gestures but also translates them into text or speech in real-time. Such innovations can contribute to better communication between the deaf and hearing communities in everyday situations.

The project could expand into educational and training applications, helping individuals learn sign language more effectively through interactive tutorials and feedback mechanisms.

REFERENCES

[1] G Ananth Rao and PVV Kishore, “Selfie video based continuous Indian sign language recognition system”. Ain Shams Engineering Journal, 2017.
[2] Swaraj Dahibavkar , Jayesh Dhopte, “Marathi Sign Language”, IJLTEMAS Volume IX, Issue V, May 2020.
[3] Becky Sue Parton, “Sign language recognition and translation: A multidiscipline approach from the field of artificial intelligence”. Journal of deaf studies and deaf education, 11(1):94–101, 2005.
[4] Y.M.Pathan, S.R.Waghmare, and P.K.Patil, “Sign Language Recognition for deaf and dumb people”, IJECS Volume 4, Issue 3, March 2015.
[5] L. Campbell, D. Becker, A. Azarbayejani, A. Bobick, and A. Pentland, “Invariant Features for 3D Gesture Recognition,” Second Int’l Conf. Face and Gesture Recognition, pp. 157-162, 1996.
[6] Rachana Patil, Vivek Patil, Abhishek Bahuguna, and Gaurav Datkhile1, “Indian Sign Language Recognition using Convolutional Neural Network”, https://doi.org/10.1051/itmconf/20214003004.
[7] J. Singha and A. Karen Das, "Recognition of Indian Sign Language in Live Video", 2013.
[8] A. Kumar Sahoo, M. Sharma, and R. Pal, "Indian sign language recognition using neural networks and KNN classifiers Image processing" View project IoT and 5G View project vol. 9, no. 8, 2014.
[9] Neelam K. Gilorkar, Manisha M. Ingle, “A Review on Feature Extraction for Indian and American Sign Language”, 2014.



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)