IJIRMET ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 10, Issue 3, March 2025

# Enhancing Communication For The Deaf Using Computer Vision To Recognize Sign Language

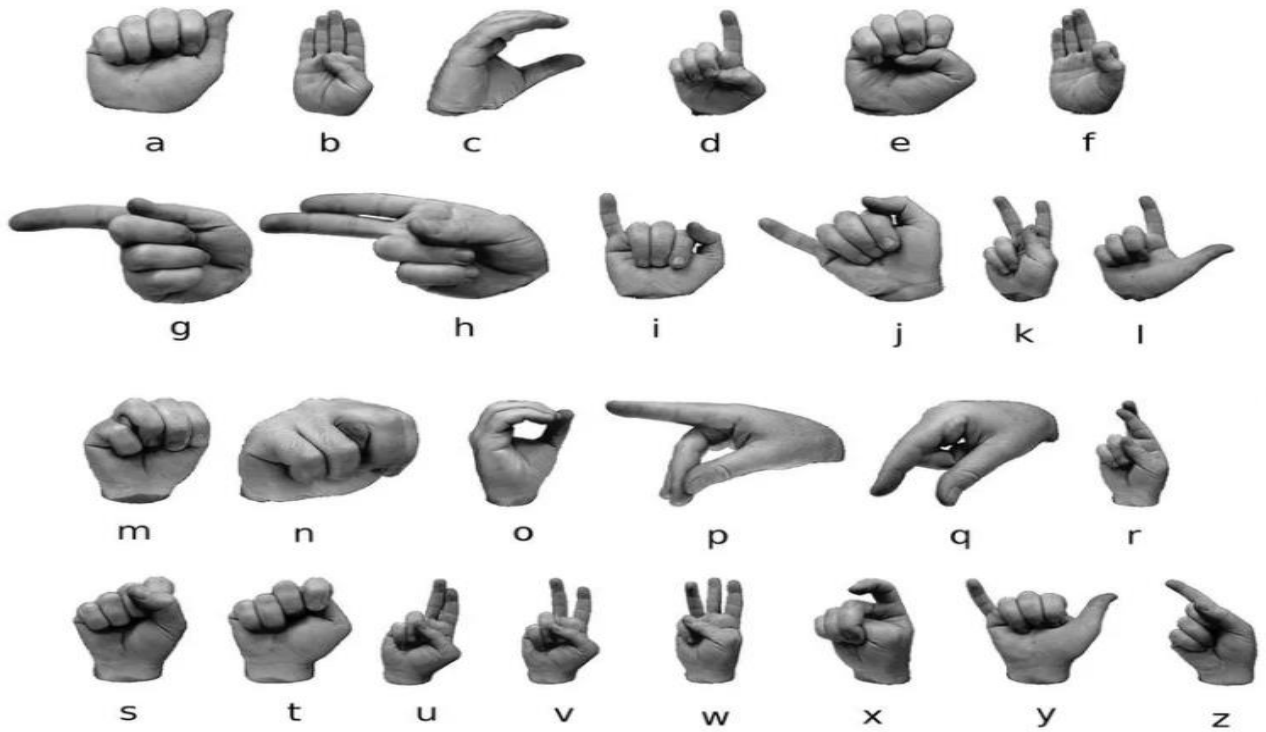
[1] S.Gopi, [2] S.Elaiyaselvi, [3] M.Kaviyadharshini, [4] P.Vishnupriya
[1] Associate Professor, [2] [3] [4] Students,Department of Information Technology, Muthayammal Engineering College (Autonomous),Rasipuram 637408,TamilNadu,India.
[1] gopi.s.it@mec.edu.in, [2] elaiyaselvis@gamil.com, [3] kaviyadharshini.m.it@gmail.com, [4] vishnupriyapr2004@gmail.com

*Abstract*: ***People with impaired speech and hearing uses Sign language as a form of communication. Disabled People use this sign language gestures as a tool of non-verbal communication to express their own emotions and thoughts to other common people. Conversing with people having a hearing disability is a major challenge. Deaf and Mute people use hand gesture sign language to communicate, hence normal people face problems in recognizing their language by signs made. Hence there is a need for systems that recognize the different signs and conveys the information to normal people. But these common people find it difficult to understand their expression, thus trained sign language expertise are needed during medical and legal appointment, educational and training session. To address this problem, we can implement artificial intelligence technology to analyse the user's hand with finger detection. In this proposed system we can design the vision-based system in real time environments. And then using deep learning algorithm named as Convolutional neural network algorithm to classify the sign and provide the label about recognized sign. Also extend the framework to capture the user speech and convert into text format using Hidden Markov Model. Finally visualize the sign-based text recognition using Natural language processing in terms of AVATARS.***

***Keywords - Hand image acquisition, Binarization, Region of finger detection, Classification of finger gestures, Sign recognition.***

## I. INTRODUCTION

The process of turning the user's signs and motions into text is referred to as sign language recognition. It helps persons who are unable to communicate with the broader public. The motion is mapped to relevant text in the training data using image processing techniques and neural networks, and so raw images/videos are turned into text that can be read and comprehended. Dumb persons are frequently denied access to normal communication with other members of society. It has been found that they find it difficult to connect with normal people with their gestures at times, as only a few of them are recognised by the majority of people. Because people with hearing loss or who are deaf are unable to communicate verbally, they must rely on some form of visual communication the majority of the time. In the deaf and dumb community, sign language is the major mode of communication. It has syntax and vocabulary much like any other language, but it communicates through visual means. The issue arises when people who are deaf or dumb try to communicate with others using these sign language grammars. This is due to the fact that most people are unaware of these grammar rules. As a result, it has been observed that a stupid person's communication is limited to his or her family or the deaf community. The increased public acceptance and funding for international projects emphasises the necessity of sign language. For the dumb community, a computer-based solution is in high demand in this age of technology. Some steps toward this goal include teaching a computer to recognise speech, facial emotions, and human gestures. Nonverbally communicated information is referred to as gestures. At any given time, a human can make an infinite number of gestures. Computer vision researchers are particularly interested in human gestures since they are received through vision. The goal of the project is to create an HCI that can detect human motions. The conversion of these motions into machine language necessitates the use of a complicated programming procedure. For better output creation, we are focused on Image Processing and Template Matching in our paper. Figure 1 depicts symbols for alphabets in sign format.




**1.2.Role of Deep Learning in Sign recognition**

Deep learning has revolutionized artificial intelligence applications, particularly in fields like computer vision and natural language processing. It enables systems to autonomously learn patterns from large datasets, significantly improving the accuracy of complex tasks such as sign language recognition. This study investigates the use of CNNs for automated recognition and real-time translation of sign language gestures.

- **Gesture and Hand Movement Recognition**: CNNs and Recurrent Neural Networks (RNNs) detect and track hand movements over time, improving accuracy.
- **Feature Extraction and Classification**: Deep learning automates feature extraction, eliminating the need for manual selection of gesture patterns.
- **Real-Time Sign Language Translation**: Models such as 3D CNNs and Transformers allow for dynamic gesture interpretation in real-time.
- **Speech and Text Conversion**: After recognition, deep learning models convert gestures into speech or text using Natural Language Processing (NLP) techniques.

## II. EXISTING METHODOLOGY

Individuals who are deaf-dumb frequently utilize sign language as a communication tool. A sign language is nothing but made of many gestures formed by distinct shapes of hand, its movements, orientations as well as the face expressions. 34 million of the 466 million individuals with hearing loss globally are children. `Deaf' people have very little or no hearing abilities. They converse with each other via sign language. Different sign languages are used by people in different parts of the world. They are extremely few in number when compared to spoken languages. Finger gesture detection efforts in the current system have been limited due to a lack of datasets and variations in sign language with locale. An ongoing project uses Indian sign language to take the first steps toward overcoming the communication gap between normal people and deaf and dumb individuals. If this initiative is successfully expanded to include words and everyday expressions, it could not only help the deaf and dumb connect with the outside world more quickly and easily, but it could also accelerate the development of autonomous systems that can comprehend and assist them. Due to a lack of standard datasets, study in Indian Sign Language lags behind that of its American counterpart.

- **Rule-Based Approaches**: Predefined rules interpret hand gestures but lack adaptability.
- **Video-Based Recognition**: Uses machine learning to analyze gesture sequences in video input.
- **Image-Based Recognition**: Trains models on static hand gesture images but struggles with real-time applications.

## III. PROPOSED METHODOLOGY

Sign Language is a gesture-based language that uses hand movements, hand orientation, and face expression in place of auditory sound patterns. This language has varied patterns based on the individual and is not universal. However, Deaf-mute people are finding it harder to communicate without the help of a translation of some kind because most people aren't familiar with sign language. They believe they are being shunned. Between deaf-mute people and normal people, Sign Language Recognition has become a commonly accepted communication approach. There are two kinds of recognition models: sensor-based systems and computer vision-based systems. In computer vision-based gesture recognition, the camera is used as an input source, and input motions are first image processed before being recognized. After that, a number of methods are employed to identify the processed gestures, including the region of interest algorithm and neural network approaches. A vision-based system for recognizing sign language has the fundamental drawback of its picture collection process being vulnerable to several environmental factors, including background circumstances, lightning sensitivity, and camera positioning. That being said, it is more affordable and practical than using a camera and tracker to gather information. However, camera data is added to neural network techniques like the Convolutional neural network for increased accuracy.

**CNN ALGORITHM:** Generally, you would take the following actions to create a CNN (Convolutional Neural Network) algorithm for sign language detection: Data Collection: Compile a sizable collection of movies or pictures with sign language. To increase the model's generalization, make sure the dataset contains a wide variety of sign gestures and variations. Data preprocessing: To improve the model's learning process, preprocess the gathered data. Resizing the photos to a consistent size, leveling the pixel values, and dividing the dataset into training and testing sets are typical preprocessing procedures. Data Enrichment: Use data augmentation methods to improve the model's generalization, such as rotating, scaling, and flipping, in order to artificially expand the dataset. This is a particularly helpful step if your original data set is small. Model Architecture: Design the architecture of your CNN model. It typically consists of convolutional layers, pooling layers, and fully connected layers. The convolutional layers extract relevant features from the input images, and the fully connected layers perform classification based on these features. Training: Train your CNN model using the prepared dataset. During training, the model learns to optimize its internal parameters by minimizing a chosen loss function, such as categorical cross-entropy, in order to make accurate predictions. Evaluation: Evaluate the trained model on the testing dataset to assess its performance. Common evaluation metrics for classification tasks include accuracy, precision, recall, and F1 score. Fine-tuning: If the model's performance is unsatisfactory, consider fine-tuning the model by adjusting the architecture, hyper parameters, or preprocessing techniques. You may also explore techniques like transfer learning, where a pre-trained model on a large dataset (e.g., ImageNet) is adapted to the sign language dataset. Deployment: Once you are satisfied with the model's performance, deploy it for real-world applications. This could involve integrating it into a mobile app or a web service that can receive input (e.g., images or videos) and provide predictions. Remember to annotate the dataset with correct labels, provide enough variations in sign gestures, and conduct rigorous testing and validation to ensure the accuracy and reliability of your CNN algorithm for detecting sign language. Figure 3 depicts the suggested framework.

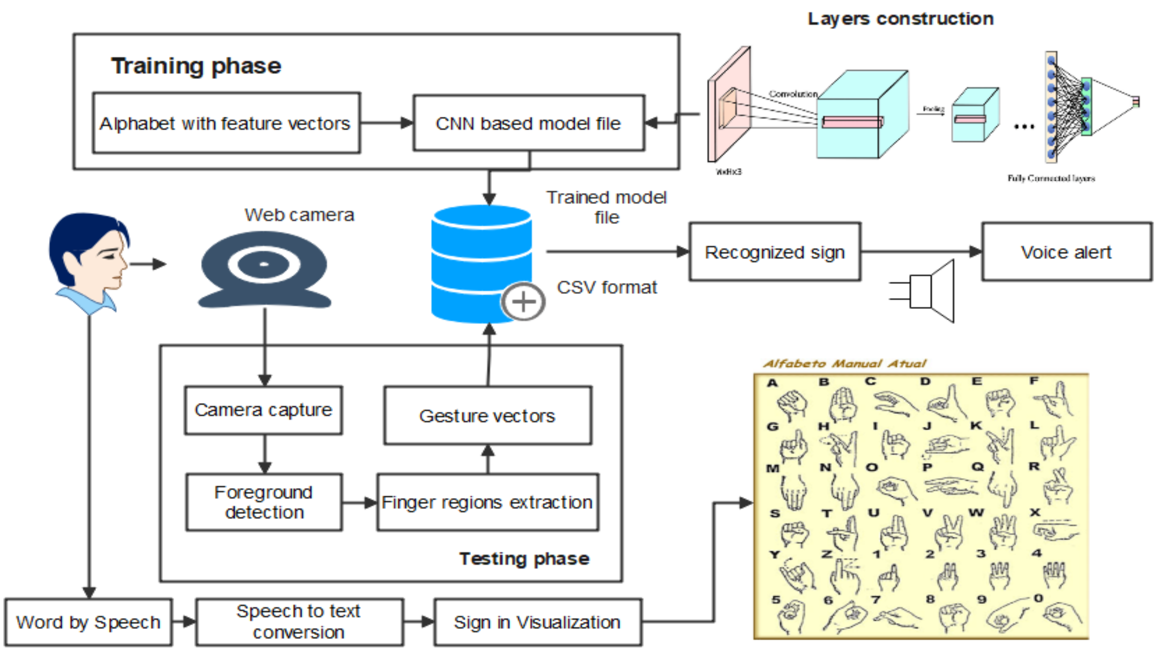



Fig 3: Proposed architecture

- **Hand Image Acquisition**: Captures hand gestures using vision-based cameras.
- **Binarization :** Employs background subtraction to isolate hand regions.
- **Finger Detection**: Extracts key features for accurate recognition.
- **CNN-Based Classification**: Processes segmented hand gestures for recognition and labeling.
- **Voice-Based Sign Visualization**: Converts recognized gestures into text and speech.

### 3.1 Advantages of Proposed Methodology

- **High Accuracy**
- **Real-Time Processing**
- **Automated Feature Extraction**
- **Scalability**
- **Robustness to Variations**
- **Multi-Modal Output**

## IV. MODULES DESCRIPTION

### 1. Hand Image Acquisition

Capturing hand gestures using a vision-based camera to obtain clear input images for processing. The system ensures high-quality input images by adjusting lighting conditions and background noise reduction. Image preprocessing techniques such as contrast adjustment and noise filtering help enhance the accuracy of the subsequent processing stages.

### 2. Binarization

Binarization converts the acquired images into a black-and-white format by applying thresholding techniques. This process helps in simplifying the image data, making it easier to extract relevant hand features while eliminating unnecessary background noise. Adaptive thresholding and edge detection methods further improve accuracy.

### 3. Region Of Finger Detection

A Region of Interest (ROI) refers to specific portions of an image that are identified for further processing. These regions contain valuable information relevant to gesture recognition. The concept of ROI is widely used in various image-processing applications to focus computational resources on key areas while ignoring unnecessary parts. In the context of sign language recognition, hand regions are extracted to isolate and analyze gestures effectively. Segmentation techniques help in distinguishing hand shapes and finger orientations. ROI ensures that only meaningful parts of the image, such as the fingers and palm, are considered, improving accuracy and efficiency in recognition. To accurately extract the hand region, **segmentation techniques** such as color-based filtering, thresholding, and edge detection are applied. These methods differentiate the hand from other parts of the image, improving precision.

**4. Classification Of Finger Gestures**

Segmentation refers to the process of partitioning a digital image into multiple meaningful segments to distinguish different objects or regions. In this context, segmentation helps in isolating the hand from the background for more accurate gesture recognition. The process involves grouping pixels into distinct categories based on shared properties such as color, intensity, or texture. This ensures that each finger gesture is effectively separated for classification. Skin color detection plays a crucial role in recognizing hand gestures. By utilizing color-based segmentation, the system can accurately differentiate hand regions from other objects in the background. Edge detection is applied to further refine the identification of finger regions by recognizing the boundaries and contours of fingers. Techniques such as Canny edge detection help highlight the edges for better gesture classification. The finger gesture classification process relies on a combination of features such as skin color, movement patterns, and edge characteristics. These attributes help in distinguishing between different gestures and improving recognition accuracy.

**5. Sign Recognition**

**Sign Recognition** is the final and most crucial phase in sign language recognition, where the extracted hand gestures are classified into meaningful signs. After the **Region of Interest (ROI) detection** and **feature extraction**, deep learning models, particularly **Convolutional Neural Networks (CNNs),** are used to analyze and categorize gestures. These models learn patterns from a large dataset of hand signs, enabling them to recognize different gestures with high accuracy. During the recognition process, the system compares the extracted features, such as **finger positions, movement trajectories, and edge contours,** against a trained model to identify the corresponding sign language symbol. Advanced models like **Recurrent Neural Networks (RNNs) and Transformers** are sometimes used to process sequential gestures, improving accuracy in recognizing dynamic signs.

**6. Voice-Based Sign Visualization**

Voice-Based Sign Visualization is the final step in the sign language recognition system, where recognized hand gestures are converted into spoken language. After the deep learning model classifies a sign, the system maps it to a corresponding word or phrase and generates an audio output using Text-to-Speech (TTS) technology. This feature bridges the communication gap between sign language users and non-sign language speakers by providing real-time verbal translation. The process involves extracting text representations of recognized gestures and feeding them into a speech synthesis model. Advanced TTS engines, such as Google's WaveNet or Microsoft’s Azure Speech, can produce natural-sounding voice outputs, enhancing accessibility. Additionally, multilingual support can be integrated to cater to different languages and dialects, making the system more inclusive. By combining deep learning for gesture recognition with speech synthesis, this module enables effective interaction in educational, healthcare, and public service settings, empowering individuals with speech or hearing impairments to communicate effortlessly.

## V. RESULTS AND DISCUSSION

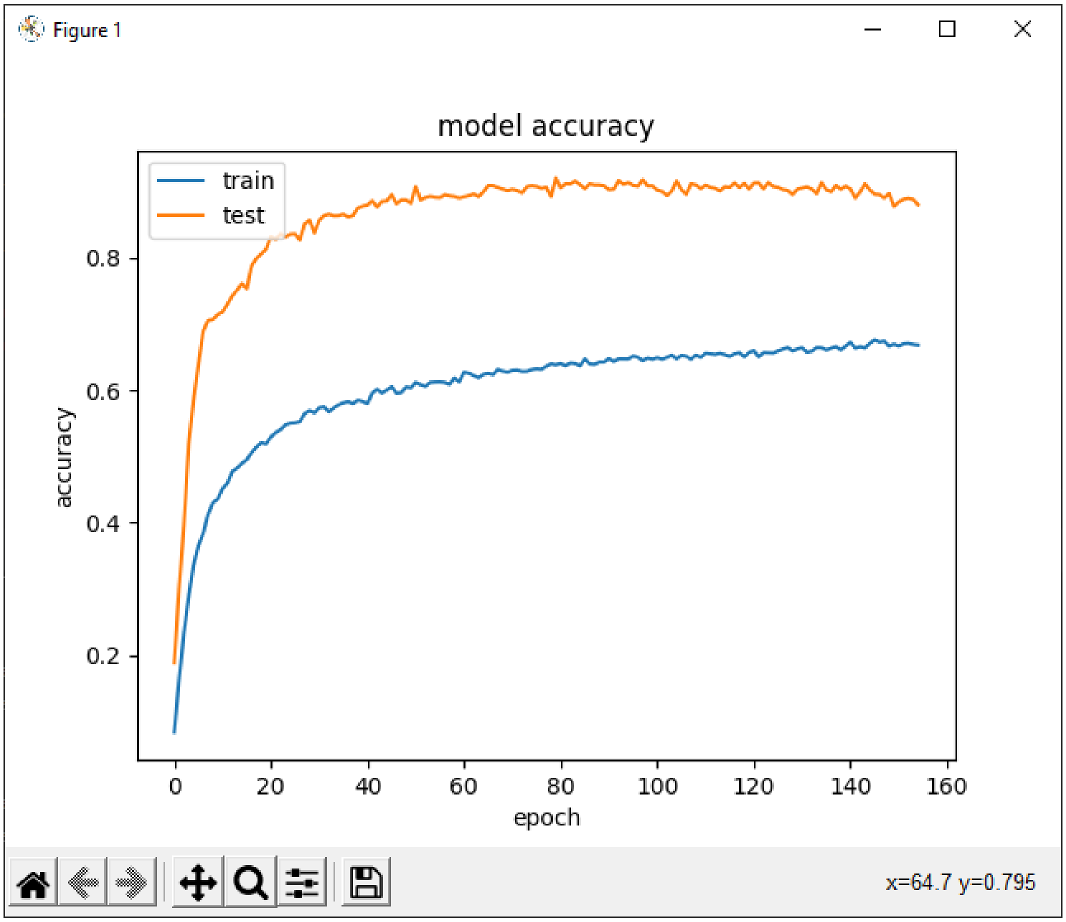
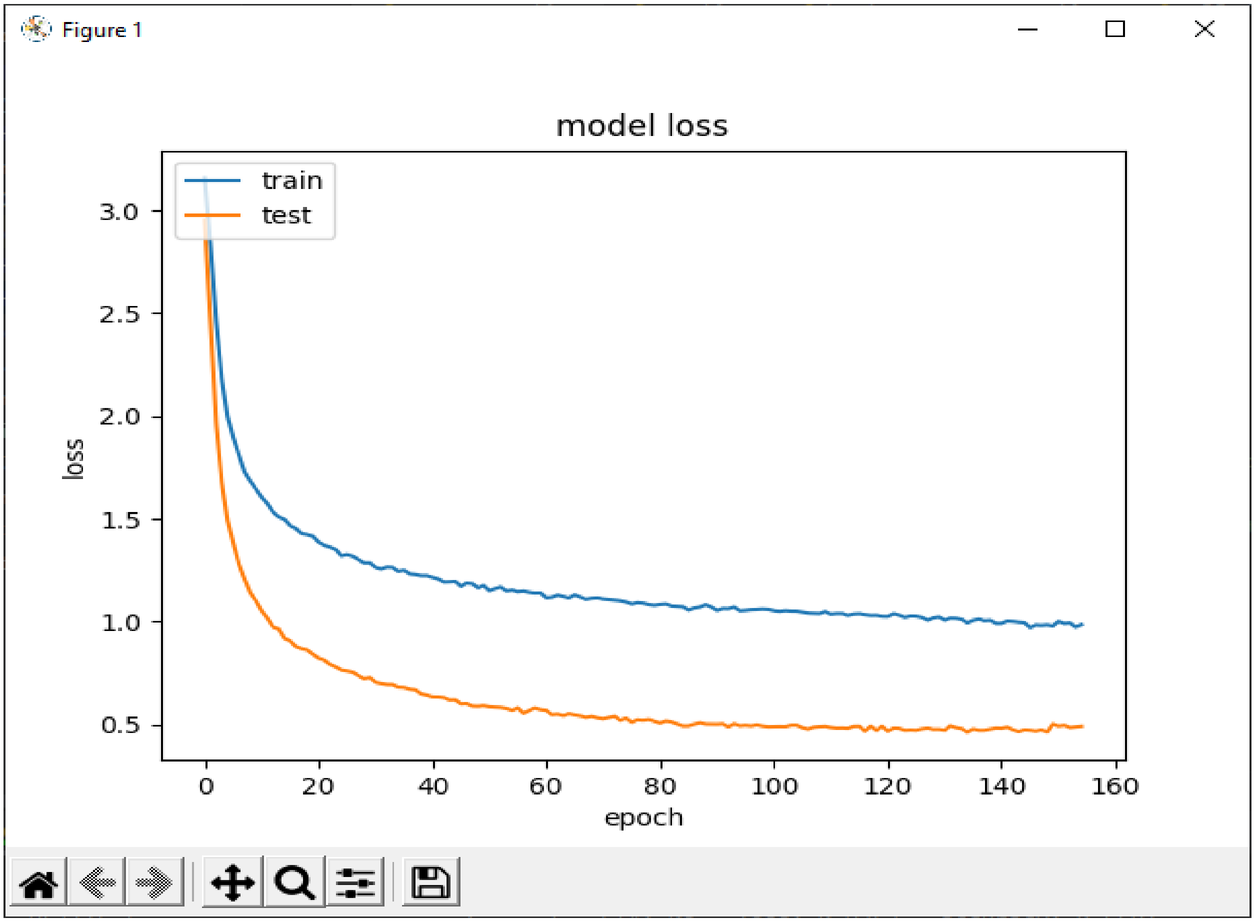


Figure 1
model accuracy
train
test
accuracy
0.2
0.4
0.6
0.8
0
20
40
60
80
100
120
140
160
epoch
x=64.7 y=0.795




**Classification Report**
Accuracy is 0.90
Macro average 0.93
Weighted average 0.9

| | Precision | recall | f-score | support |
|---|---|---|---|---|
| 0 | 0.99 | 0.90 | 0.95 | 105 |
| 1 | 0.91 | 1.00 | 0.95 | 99 |
| 2 | 0.92 | 0.66 | 0.77 | 146 |
| 3 | 0.73 | 0.58 | 0.65 | 185 |
| 4 | 0.63 | 0.95 | 0.76 | 168 |
| 5 | 1.00 | 0.85 | 0.92 | 88 |
| 6 | 1.00 | 0.92 | 0.96 | 39 |
| 7 | 0.98 | 0.75 | 0.85 | 73 |
| 8 | 0.91 | 1.00 | 0.95 | 86 |
| 9 | 0.96 | 1.00 | 0.98 | 54 |
| 10 | 0.93 | 0.93 | 0.93 | 123 |
| 11 | 0.95 | 0.83 | 0.89 | 196 |
| 12 | 1.00 | 0.97 | 0.98 | 230 |
| 13 | 0.97 | 1.00 | 0.98 | 143 |
| 14 | 0.95 | 0.97 | 0.96 | 125 |
| 15 | 0.99 | 1.00 | 0.99 | 228 |
| 16 | 0.99 | 0.96 | 0.97 | 193 |
| 17 | 0.87 | 1.00 | 0.93 | 256 |
| 18 | 0.95 | 0.99 | 0.97 | 117 |

| | | | | |
|---|---|---|---|---|
| 19 | 0.64 | 1.00 | 0.78 | 140 |
| 20 | 0.92 | 0.87 | 0.89 | 165 |
| 21 | 1.00 | 0.98 | 0.99 | 134 |
| 22 | 0.99 | 1.00 | 1.00 | 139 |
| 23 | 0.99 | 0.54 | 0.70 | 156 |
| 24 | 0.99 | 1.00 | 0.99 | 96 |
| 25 | 0.94 | 0.95 | 0.95 | 168 |

## VI. CONCLUSION AND FUTURE ENHANCEMENT

Seeing, hearing, speaking, and reacting to situations correctly are some of the most precious gifts that a person can possess. However, some unfortunate people are denied this opportunity. Sharing ideas, thoughts, and experiences with those around them helps people come to know one another. There are a few methods to do this, the most effective being the ability to "Speech." Everyone is able to communicate through speech in a very persuasive way and understand one another. Our initiative intends to close the gap by including a low-cost computer into the communication chain, allowing sign language to be captured, recognised, and translated into speech for the benefit of blind individuals. This paper uses an image processing technique to identify the handcrafted movements. A contemporary integrated designed system for those with hearing impairments is shown through this application. The user's data collection can be facilitated by the camera-based zone of interest. Every single deed will have significance on its own.

**Future Enhancement**

Future improvements can focus on expanding the system to support dynamic sign language recognition, allowing for the interpretation of continuous gestures rather than isolated signs. The integration of Natural Language Processing (NLP) could enhance contextual understanding, improving the accuracy of sentence formation. Additionally, incorporating 3D hand tracking, pose estimation, and multimodal learning could further refine gesture recognition. Real-world deployment through mobile applications and wearable devices would make the system more accessible, ensuring seamless communication for users in various environments.

## REFERENCES

[1] Xu, Huanyuan, et al. "RF-CSign: A Chinese Sign Language Recognition System Based on Large Kernel Convolution and Normalization-Based Attention." IEEE Access 11 (2023): 133767-133780.
[2] Zhang, Menglin, Shuying Yang, and Min Zhao. "Deep Learning–based Standard Sign Language Discrimination." IEEE Access (2023).
[3] Abdullahi, Sunusi Bala, and Kosin Chamnongthai. "IDF-Sign: Addressing Inconsistent Depth Features for Dynamic Sign Word Recognition." IEEE Access (2023).
[4] Jafari, Farzaneh, and Anup Basu. "Two-Dimensional Parallel Spatio-Temporal Pyramid Pooling for Hand Gesture Recognition." IEEE Access 11 (2023): 133755-133766.
[5] Ansar, Hira, et al. "Hand Gesture Recognition for Characters Understanding Using Convex Hull Landmarks and Geometric Features." IEEE Access (2023).
[6] Gu, Yutong, Hiromasa Oku, and Masahiro Todoh. "American Sign Language Recognition and Translation Using Perception Neuron Wearable Inertial Motion Capture System." *Sensors* 24.2 (2024): 453.
[7] Teran-Quezada, Alvaro A., et al. "Sign-to-Text Translation from Panamanian Sign Language to Spanish in Continuous Capture Mode with Deep Neural Networks." *Big Data and Cognitive Computing* 8.3 (2024): 25.
[8] Renjith, S., Manazhy Rashmi, and Sumi Suresh. "Sign language recognition by using spatio-temporal features." *Procedia Computer Science* 233 (2024): 353-362.
[9] Arooj, Sadia, et al. "Enhancing sign language recognition using CNN and SIFT: A case study on Pakistan sign language." *Journal of King Saud University-Computer and Information Sciences* 36.2 (2024): 101934.
[10] Al Khuzayem, Lama, et al. "Efhamni: A Deep Learning-Based Saudi Sign Language Recognition Application." *Sensors* 24.10 (2024): 3112.