SIGN LANGUAGE RECOGNITION SYSTEM

PALLAPOTHU VENKATA SAI VARDHAN
Department of Computer Science Engineering
Parul Institude of Engineering and Technology
Vadodara, Gujarat, India
200303128013@paruluniversity.ac.in

YALLURI OMKARNATH
Department of Computer Science Engineering
Parul Institude of Engineering and Technology
Vadodara, Gujarat, India
200303126197@paruluniversity.ac.in

VALIVETI UDDAV RISHIK
Department of Computer Science Engineering
Parul Institude of Engineering and Technology
Vadodara, Gujarat, India
200303126191@paruluniversity.ac.in

VURA SRI CHANDAN
Department of Computer Science Engineering
Parul Institude of Engineering and Technology
Vadodara, Gujarat, India
200303126196@paruluniversity.ac.in

KISHORI SHEKOKAR
Project Mentor
Department of Computer Science Engineering
Parul Institude of Engineering and Technology
Vadodara, Gujarat, India
kishori.shekokar20174@paruluniversity.ac.in

***Abstract*—The sign language recognition system is an innovative device that overcomes the communication gap between the hearing and deaf cultures. It enables real-time translation of sign language into spoken language and vice versa, promoting inclusivity and accessibility in our increasingly diverse society. Using advanced computer vision and machine learning algorithms, the system analyzes the movements of a hands, facial expressions, and body language to accurately recognize and interpret the meaning of their signs. It can recognize a wide range of sign language styles, and adapt to the individual habits of different users. The system utilizes a camera to capture sign language gestures and recognizes them in real-time. The proposed system comprises of three main stages, hand detection and tracking, feature extraction, and classification. In the hand detection and tracking stage, a hand region of interest is obtained and tracked using a skin color-based method. In the feature extraction stage, hand shape and motion features are extracted using geometric and temporal features, respectively. A support vector machine (SVM) is used to recognize sign language motions during the classification step.**

With its user-friendly interface and intuitive design, the sign language recognition system can be easily integrated into various devices and applications, such as smartphones, laptops, and smart home assistants. It opens up new possibilities for deaf and hard-of-hearing individuals in education, employment, and social interaction, enabling them to communicate more effectively and confidently in a hearing-dominated world. It is a ground-breaking technology that empowers the deaf and hard-of-hearing communities and promotes diversity and inclusion in our society.

***Index Terms*—Sign language, Machine Learning,Image Processing SVM, CNN, RNN, Mediapipe, NLP, Text to Speech.**

I. INTRODUCTION

A device known as a sign language recognition system enables computers to decipher and comprehend the motions and movements used in sign language. It is a crucial tool for those who are hard of hearing or deaf because it gives them a way to communicate with others who are not proficient in sign language. The system recognizes and converts sign language into text or voice using computer vision and machine learning algorithms. A significant advancement in sign language recognition technology has made it possible for computers to understand and decode the complex language of gestures and motions. This invention facilitates smooth communication between those who are not familiar with sign language and those who do.

At its essence, the system harnesses the power of computer vision and machine learning. Computer vision captures and dissects the subtleties of hand shapes, motions, facial expressions, and body postures inherent to sign language. Machine learning models, trained on extensive sign language data sets , then transform these visual cues into meaningful text or speech.

The Sign Language Recognition System is lifeline for those reliant on sign language for communication. From educational tools to healthcare and customer service, the applications of sign language recognition are far-reaching. Moreover, it addresses persistent challenges in the field and glimpses into the future, where these systems continue to enhance accessibility and communication in an interconnected world.

A device called an American Sign Language (ASL) recognition system is intended to help those who may not be competent in sign language communicate with the deaf and hard of hearing community. This system translates sign language movements into spoken language or text by using cameras, sensors, machine learning, and computer vision. Using cameras or wearable technology, the signer's hand and facial expressions are first recorded. Then, using computer vision algorithms, these motions are detected and tracked while important information like hand shape, orientation, and face expressions are extracted. To detect and understand the signs, machine learning models—which are frequently built on deep learning—are trained on enormous databases of sign language.

The technology is accessible to deaf people and others they are talking with since it offers feedback through user-friendly interfaces. Adapting different signing techniques and geographical differences in sign language are challenges. Applications promise to improve inclusivity and accessibility in a variety of situations, from instructional tools for sign language learners to communication assistance for the deaf.

II. METHODOLOGY

A organized set of stages is involved in the technique for developing a sign language recognition system. First, a large and varied collection of sign language expressions and motions with a variety of variants must be gathered. The next step is data preparation, which involves ensuring consistency by cleaning, normalizing, and aligning the data.Feature extraction is a vital stage since it involves extracting key characteristics such as hand form, direction, movement, and facial expressions. When deciding on the best machine learning or deep learning model, convolutional neural networks (CNNs), recurrent neural networks (RNNs), or advanced models such as transformers are commonly recommended. With validation and testing datasets, the model's performance is assessed after it has been trained on the preprocessed data.

Cameras, depth sensors, or wearable technology are used to enable real-time gesture recognition, and computer vision algorithms are used to follow and analyze hand and facial motions. Real-time classification and interpretation of the facial expressions and sign language motions by the trained model translates them into text or spoken language. The user interface of the system need to be easily navigable and intuitive, offering both the signer and the recipient feedback. The system can learn and adjust to different signing styles with the help of customization and adaptation, which are crucial. Crucial components of the process include integration into several platforms, scalability, cross-validation with multiple signers, and continuous feedback loops. Cooperation with the community of hard-of-hearing people is necessary to guarantee cultural sensitivity and fit with their requirements.

Furthermore, constant improvements and modifications based on user input and technology breakthroughs are essential for enhancing the accuracy, flexibility, and inclusiveness of the system while taking ethical issues like data privacy and accessibility into account.

A. Dataset

The American Sign Language dataset, or ASL dataset. This dataset comprises a substantial collection of 31,000 images, thoughtfully distributed with 1,000 images allocated to each of the 31 distinct classes. These classes represent a rich variety of sign language gestures, all of which were meticulously documented with contributions from five different subjects.

The gestures featured in this dataset encompass numerals ranging from the full set of alphabets from 'A' to 'Z', with the exception of 'J' and 'Z'. This omission was necessary as 'J' and 'Z' involve dynamic hand movements, making it challenging to represent them accurately within the confines of static images.

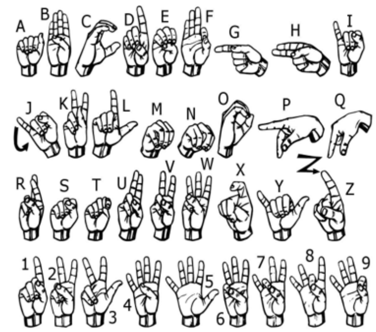

Fig. 1. ASL Dataset.

The process of creating a dataset involves setting goals in a methodical manner, gathering information from several sources via web scraping or APIs, and organizing it into an organized style. Data preparation, which includes feature engineering, normalization, and cleaning, is crucial. If necessary, data annotation involves classifying or labeling data. Version control, data augmentation, and data separation all support the administration and quality of datasets.

B. Mediapipe detection

The MediaPipe framework is designed for creating machine learning pipelines for analyzing time-series data such as audio and video. In 2019, it was made freely available so that scholars and developers may include it into their projects.The ability of MediaPipe to function accurately and robustly on low-powered platforms, such as Android and IoT devices, is what makes it stand out.

Systems that recognize sign language rely heavily on this adaptable computer vision architecture. It offers a reliable platform for monitoring and evaluating the body motions, face expressions, and hand movements that are crucial to sign language communication. MediaPipe improves the accuracy and efficiency of sign language recognition systems by streamlining the process of extracting important sign language elements through its pre-trained models and simple integration. Its capacity to comprehend and translate sign language motions in real time makes it an effective tool for improving communication among the deaf and hard of hearing.

Mediapipe's significance in this industry stems from a number of its essential qualities, which include:

Hand tracking is a crucial skill for understanding the complex hand gestures and forms that are a fundamental part of sign language. The real-time detection and tracking of hands

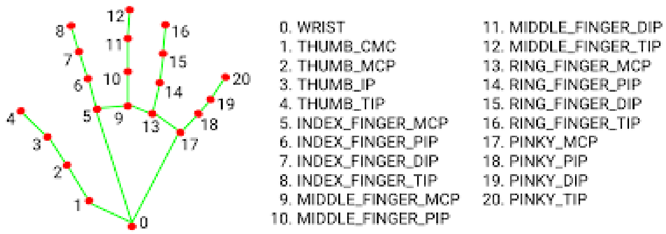

Fig. 2. Gesture Recognition.

using Mediapipe's hand tracking module allows for accurate interpretation of sign motions.

Pose Estimation: When interpreting sign language, it's crucial to identify the signer's body language as well as their face emotions. Pose estimation features in Mediapipe can aid in capturing these non-manual inputs.

Real-time Processing: To enable smooth communication between signers and non-signers, real-time processing is frequently needed for sign language recognition. High-performance recognition and minimal latency are guaranteed by Mediapipe's effective processing.

Customization: Mediapipe may be tailored to different sign languages and unique recognition requirements by allowing developers to build bespoke pipelines and add their own models.

Accessibility: Developers and academics may work on sign language recognition projects with more ease because to Mediapipe's strong and open-source foundation, which eventually helps the Deaf and hard-of-hearing population.

Fig. 3. Hand image .

Fig. 4. Hand image .

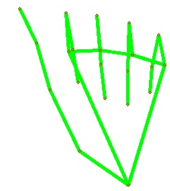

Fig. 5. mediapipe image 1.

C. Machine Learning Modules

We used a variety of machine learning classifiers, such as Image processing,SVM ,KNN,CNN,NLP,Text to speech.

Image Processing

By doing the image processing in sign language recognition , we can do firstly image acquisition and after that image pre-processing.

Image acquisition

It is critical in a system that recognizes sign language since it is the basis for the system's capacity for understanding and interpret sign language motions. Image acquisition involves capturing visual information from the signing person, typically through cameras or depth sensors, and transforming this visual input into digital data that can be processed by the system.

Image acquisition provides the system with the raw visual data it needs to analyze and recognize sign language gestures. This information comprises hand forms, actions, movements of the face, and body postures, all of which are necessary for sign language communication.

Image pre-processing

Before any recognition can take place, the system must first process the input video or image to enhance contrast, remove noise, and normalize the image size and orientation.

Preprocessing a picture usually involves a few important stages. These comprise of normalization to guarantee uniform scale and orientation, picture segmentation to separate the pertinent elements like hands and facial expressions, and noise reduction to get rid of undesired visual distortions. Furthermore, methods for feature extraction are commonly employed to identify and quantify relevant visual characteristics, and color correction and backdrop removal can be used to enhance the appearance of sign language components.

SVM

Support Vector Machine (SVM) represents a popular supervised teaching technique that may be used for regression as well as classification applications. However, it is most commonly used for problems with classification in machine learning.

The SVM algorithm's goal is to determine the ideal decision boundary or line that can divide a space with n dimensions into classes, permitting us to easily place new data points in the appropriate category in the not-too-distant future. A hyperplane is the optimal option boundary.

CNN

Convolution-Based Neural Systems—most notably, Convolutional Neural Networks (CNNs)—are essential to systems that recognize sign language because they help analyze body language, facial expressions, and hand motions. CNNs are ideally suited for the complexities of sign language since they are highly skilled at obtaining spatial characteristics from video frames. The hierarchical qualities that are necessary for identifying the intricate hand forms, postures, and motions that take place during signing may be captured by them. These networks are also useful for simulating dynamic sign language expressions because they can adjust to both geographical and temporal input.

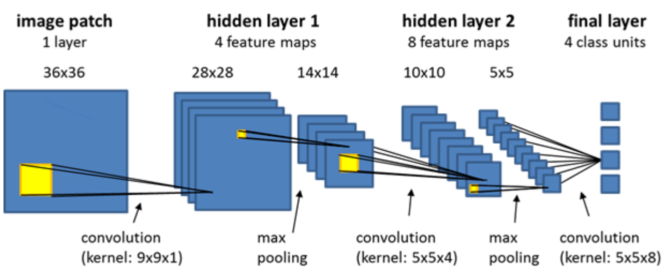

Fig. 6. CNN.

RNN

n sign language recognition systems, recurrent neural networks (RNNs) are essential for improving the temporal modeling part of the recognition process. Because sign language involves dynamic hand and body motions that are intrinsically temporal, RNNs are particularly engineered to capture these sequential patterns. They are particularly good at managing long-term dependencies in sign language, which is essential for effective interpretation, and identifying the order in which signals and gestures should be used. They also grasp the context of a sentence or phrase. In order to provide coherent and contextually appropriate translations, RNNs are frequently employed to translate detected signs into text or spoken language.

NLP

The use of natural language processing, or NLP, in sign language recognition systems aims to increase accessible and interaction for the deaf and hard of hearing people. When it comes to captioning movies or converting sign language movements into printed text, natural language processing (NLP) approaches play a crucial role in making the content accessible to a broader audience. Furthermore, NLP and text-to-speech (TTS) technology may be used to translate sign language into spoken language, allowing signers and non-signers to communicate more effectively. Natural language processing (NLP) models facilitate the production of natural sign language interpretations from spoken or written language by helping to comprehend the context and intent behind sign language messages.

Text to Speech

The fundamental aim of a sign language recognition system is to facilitate communication between signers and non-signers by interpreting and translating sign language gestures into written or spoken language. Text-to-speech, or TTS, recognition is usually not the main feature of these systems. Rather, the system focuses mostly on gesture detection, image processing, computer vision, and feature extraction. In order to categorize and detect the motions, the procedure entails obtaining video recordings or image frames of sign language gestures, preprocessing the data to improve its quality, extracting pertinent elements including hand forms and movements, and using machine learning techniques.

III. Classification

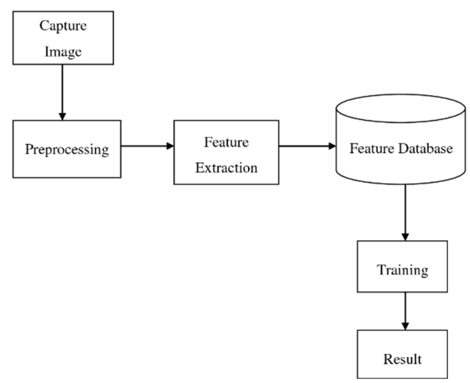

Fig. 7. Classification.

A. *Stages of execution*

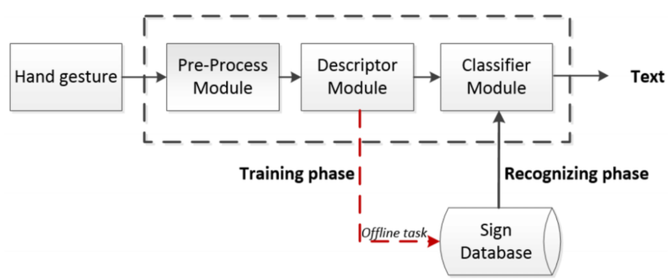

Fig. 8. SLR flowchart.

Hand tracking

The suggested method starts with hand tracking and detection. Finding and following the hand area of interest (ROI) in the incoming video stream is the aim of this step. In our technology, we separate the hand region from the backdrop using a skin color-based technique.

We extract the RGB ROI from the entire image and transform it to a grayscale image, as seen below.

Fig. 9. .Greyscale

Lastly, we add a gaussian blur filter to our image, which aids in the extraction of different characteristics. After adding gaussian blur, the picture appears as follows.

Fig. 10. Guassian blur image.

Feature Extraction

The second stage of the proposed system is featuring extraction. The purpose of this stage is to extract hand shape and motion features from the tracked hand region. We use geometric features to extract the hand shape, which includes the hand width, height, and aspect ratio. To extract the hand motion features, we use temporal features, which capture the motion information of the hand over time. Specifically, we use optical flow to estimate the motion of each pixel in the hand region, and then calculate the average motion magnitude and orientation as the hand motion features.

We have collected the different signs and angles for sign letter A to Z.

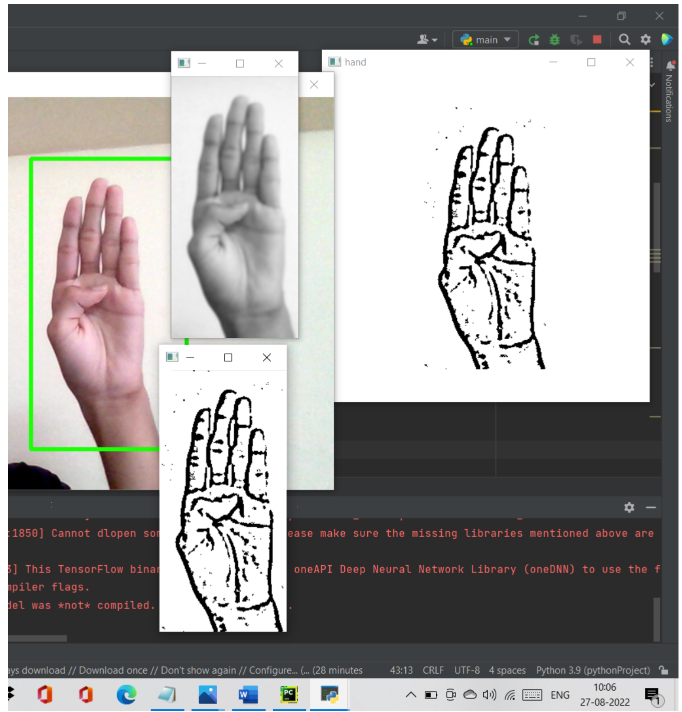

Fig. 11. Pre-processing Image.

By acquiring these landmark points and rendering them on a blank, white background using the Open CV library, we effectively address the challenges posed by varying background and lighting conditions. The utilization of the Mediapipe library to extract landmark points ensures that we can reliably obtain hand gestures in diverse background settings and under different lighting conditions.

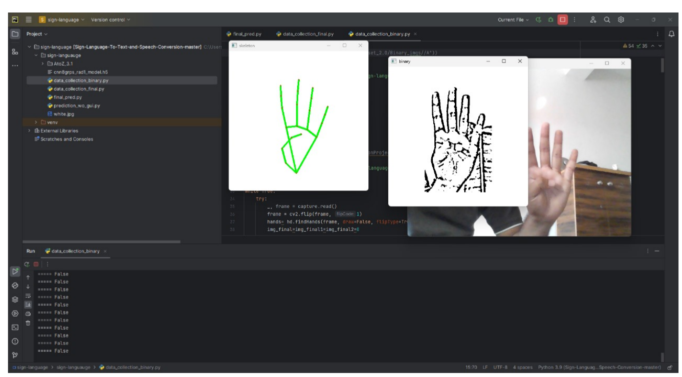

Fig. 12.

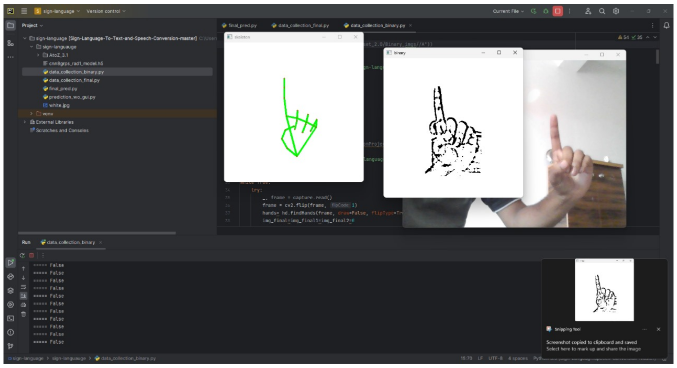

Fig. 13.

B. Classification Report

Here we can classify the sign language recognition process and its predict the letter by using the signs.

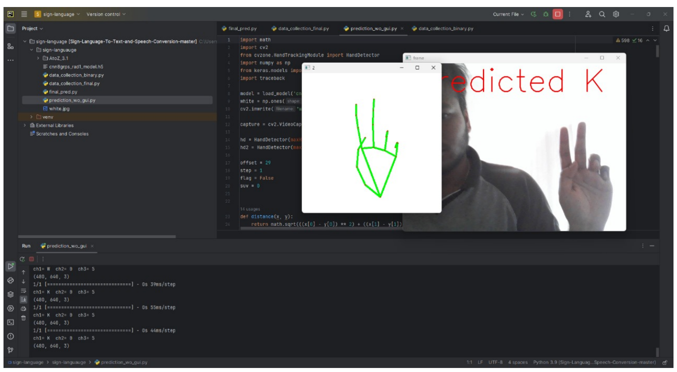

Fig. 14. Predict 1.

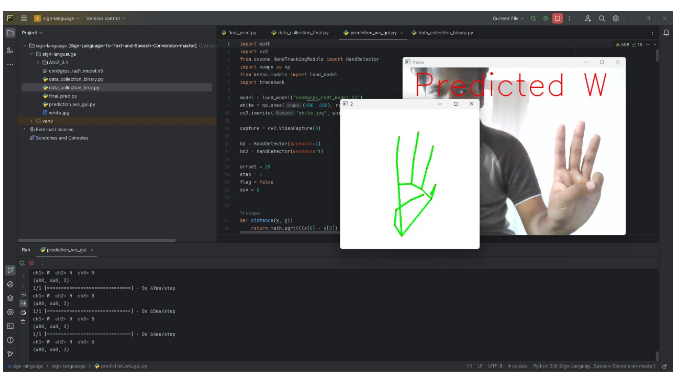

Fig. 15. Predict 2.

Here is the Performance evaluation table of symbols in dataset with its rate of recognition:

Symbol	Rate of recognition	Symbol	Rate of recognition
A	95	N	80
B	90	O	85
C	90	P	90
D	90	Q	80
E	90	R	90
F	90	S	85
G	90	T	90
H	90	U	90
I	90	V	95
J	90	W	90
K	90	X	75
L	90	Y	90
M	80	Z	85

TABLE I
AN EVALUATION TABLE.

The output of the process after all the testing and training the images done .

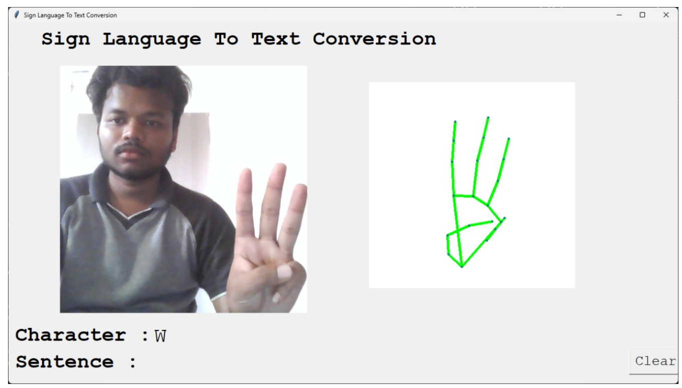

Fig. 16. OUTPUT 1.

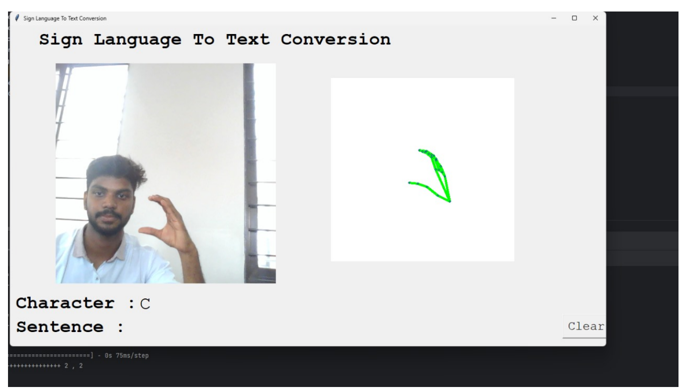

Fig. 17. OUTPUT 2.

IV. Conclusion

Systems for recognizing sign language have the potential to greatly improve accessibility and communication for the deaf and hard of hearing. These systems use a range of technologies to identify and interpret sign language motions, such as sensor-based devices, computer vision, and machine learning.

But when these systems are developed, there are still some issues that need to be resolved. For instance, distinct sign languages are used in various areas and civilizations, proving that sign language is not universal. Also, the sign language is a complex visual syntax that integrates body language, face gestures, and hand movements.

The development of sign language recognition systems has advanced significantly in recent years, despite these obstacles. With greater study and development, these systems may improve in accuracy, dependability, and accessibility, benefiting persons who are deaf or hard of hearing by allowing for improved communication and improving their overall quality of life.

As we look toward the future, the potential of sign language recognition remains boundless. Its integration with augmented reality, wearable devices, and its adoption in everyday applications promises even greater accessibility and inclusivity.

In summary, the Sign Language Recognition System is a ray of hope that dismantles obstacles to communication and promises a better, more accepting future for those who express themselves via sign language.

V. Future Work

The future of sign language recognition is bright, and there are several areas where further research and development are needed to improve the performance and effectiveness of these systems. Here are some potential future works for sign language recognition:

- Improved Accuracy
- Real time Translation
- Adaption to regional Differences
- Recognition of Facial Expressions
- Integration with other Technologies

References

[1] "Real-time Assamese sign language recognition using media pipe and deep learning" by Jyotishman Bora , Saine Dehingia ,Published in science direct on 2023.

[2] "A vision-based deep learning approach for independent users Arabic sign language interpretation" published in springer link on 2023.

[3] " CVT-SLR: Contrastive Visual-Textual Transformation for Sign Language Recognition with Variational Alignment" published in IEEE Xplore on 2023.

[4] " A comparative study of evaluating and bench marking sign language recognition system-based wearable sensory devices using a single fuzzy set" published in science direct on 2023.

[5] "Sign language identification and recognition" From the journal Open Computer Science published by De Gruyter open Access on 2022.

[6] "Sign Language Recognition: Different Types, Modalities, and Datasets" by M. Madhiarasan, Published on April 2022.

[7] "Sign language recognition in context of vision-based and deep learning" by S.Murugavalli, Published in science direct on 15 July 2022.

[8] " OpenHands: Making Sign Language Recognition Accessible with Pose-based Pre-trained Models across Languages" published on 2022.

[9] " Indian Sign Language recognition system using SURF with SVM and CNN" published in science direct on 2022.

[10] " Sign Pose-based Transformer for Word-level Sign Language Recognition" published in IEEE on 2022.

[11] "Sign Language Recognition for Deaf and Dumb" by R Rumana, R Prema, published in 2021.

[12] "Sign Language Recognition Using Media pipe Framework with Python" by Adarsh Vishwakarma; Niraj Yadav; Shree L.R., Published in IJSRD on 1ST JUNE 2021.

[13] " Machine learning methods for sign language recognition: A critical review and analysis"published in science direct on 2021.

[14] " American sign language recognition and training method with recurrent neural network" published in science direct on 2021.

[15] " Skeleton Aware Multi-modal Sign Language Recognition" published in IEEE on 2021.

[16] "ML Based Sign Language Recognition System" by K. Amrutha; P. Prabhu, Published on IEEE in 2021.

[17] "Real Time Sign Language Recognition and Speech Generation" by Amrita Thakur, Published in JIIP on 2020.

[18] "Vision-based Portuguese Sign Language Recognition System" Paulo Trigueiros and Luís Paulo Reis, Published in JIIP on 2020.

[19] "Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective" Published in JIIP on 2020.

[20] "ANFIS Based Methodology for Sign Language Recognition and Translating to Number in Kannada Language" by Ramesh M. Kagalkar and Dr. S.V Gumaste, Published in iJES on 2017.

[21] "Sign Language Recognition using Deep Learning Techniques" by P. Gnanasekar, K. Srinivasan, and K. Karthikeyan, published by Springer International Publishing in 2019, is a relevant book for this topic.

[22] "Sign Language Recognition Using Neural Networks" by Y. Jung and K. Lee: This book covers various neural network models and architectures for sign language recognition.

[23] "Pattern Recognition and Machine Learning" by Christopher M. Bishop.

[24] "Sign Language Recognition, Translation and Production: A Multimodal Information Access Perspective" by David Filiberto .

[25] "Sign Language Recognition and Translation: A Multidisciplinary Approach" by M. M. Srinivasan, S. S. Narayanan, and A. K. Jain.

[26] "Sign Language Recognition: A Machine Learning Approach" by Manu Kumar.

[27] "Gesture and Sign Language in Human-Computer Interaction" by Ewa Monika Sibinska and Pawel Rutkowski .

[28]

[29] "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

[30] "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition" by Daniel Jurafsky and James H. Martin.

[31] "Sign Language Recognition with Microsoft Kinect" by Pradeep Pasupuleti and Sudha Srinivasan.