Visual Language Interpreter

Aniket Jadhav[1]; Tejas Ulawekar[2]; Shubham Kondhare[3]; Nirbhay Mokal[4]; Rupali Patil[5]

[1-5]Dept. of Computer Engineering Pillai HOC College of Engineering and Technology (University of Mumbai)

Publication Date: 2025/03/27

Abstract: Meaningful communication is a basic human need, and yet there are some people who make use of sign language to communicate with the spoken word and encounter serious obstacles. This disconnect can leave us feeling isolated and alienated. Our project aims to solve this issue by creating a system which recognizes few hand signs that real-time converts into spoken as well written text. Our aim is to create a solution that can enable efficient natural language processing and an efficient gesture recognition, which will be based on convolutional Networks (CNN) and deep learning technology. Our text prediction: It improves the translation provided in terms of accuracy and relevance, as well as shortens processing time and communication. CNNs are a type of deep models, and designed to process structured data represented in form of 2D grids or multiarray like digital images. They operate by extracting and understanding features from visual inputs, using a hierarchy of filters that automatically recognize different patterns at increasing levels. Sign language is a critical example of the nuanced gestures these features would enable us to better understand. Our system then generates and can identify these different hand movements quite accurately. This enables these same gestures to be translated effortlessly into both speech and text, thus improving communication for sign language dependent persons. In addition, our solution consists of leading-edge text prediction technologies for optimization in translation. The purpose of these algorithms — increasing the accuracy and relevance of translations while at the same time decreasing both processing times, rendering communication quicker and more natural.

Keywords: *Sign Language, Convolutional Neural Networks (CNN), Deep Learning, Gesture Recognition, Text Prediction, Machine Learning, Artificial Intelligence***.**

How to Cite: Aniket Jadhav; Tejas Ulawekar; Shubham Kondhare; Nirbhay Mokal; Rupali Patil (2025). Visual Language Interpreter. *International Journal of Innovative Science and Research Technology,* 10(3), 1085-1091. https://doi.org/10.38124/ijisrt/25mar920

I. INTRODUCTION

Sign language (SL) is an essential communication tool for millions of individuals around the globe, according to World Health Organization, the individuals with hearing disabilities also prefer using sign language, an estimated 430 million people worldwide are completely deaf, while 1.5 billion individuals are affected by partial hearing impairment. Despite its importance, SL has historically been overlooked from research perspective when been in a comparison with the spoken languages due to its unique and differential structural characteristics. This difference leads to unique considerations as far as effective translation and recognition as well as compelling needs for solutions. Abdullah Al. et al. Cheng et al.'s [28] assessment of 58 research papers on SL translation highlights the implementations of deep learning methods, including convolutional and recurrent neural networks, to improve recognition accuracy. This paper proposes the solution of integrating manual and non-manual features (face and body language) into SL recognition systems and proves that manual and non-manual features do contribute significantly to improvement of SL recognition systems. But we are facing challenges, such as dynamic sign complexity or robust features extractor. The authors recommend further examination of deep learning models and applying the systems to real-world systems.

II. BACKGROUND

Sign translation is one of the most anticipated study areas, emphasizing the necessity of creating a communication channel between hearing and non-hearing cultures. In an effort to close the gap between the hearing challenged and the general population, the last few decades has seen a lot of research in this field. It lies at the intersection of artificial intelligence, linguistics, machine learning, computer vision, and human-computer interaction. The highest difficulty in this area is to create reliable systems that allow translation of visual, kinetic and mouthing components of sign language into typed. Many techniques, methodologies, and technologies have been previously investigated. In the early days, researchers concentrated on building still systems capable of recognizing a small number of gestures with hardware-heavy approaches. So-called data gloves with sensors were commonly deployed to trace hand movements and placements. For example, the "Power Glove" project in

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar920

the 1990s made a pioneering attempt at recognizing static hand gestures. However, such systems were limited by their reliance on specific hardware and were non-scalable. The incorporation of machine learning was a monumental shift in this space. SIFT (Scale-Invariant Feature Transform), HOG (Histogram of Oriented Gradients) etc. were the first feature extraction methods that researchers used with hand gesture images for extracting the representations from such images. These movement and stance patterns were converted into sign language words or phrases using classification models like Support Vector Machines (SVMs) and Hidden Markov Models (HMMs). However, such systems under-performed in terms of real-time efficiency and failed to generalize well on different datasets.

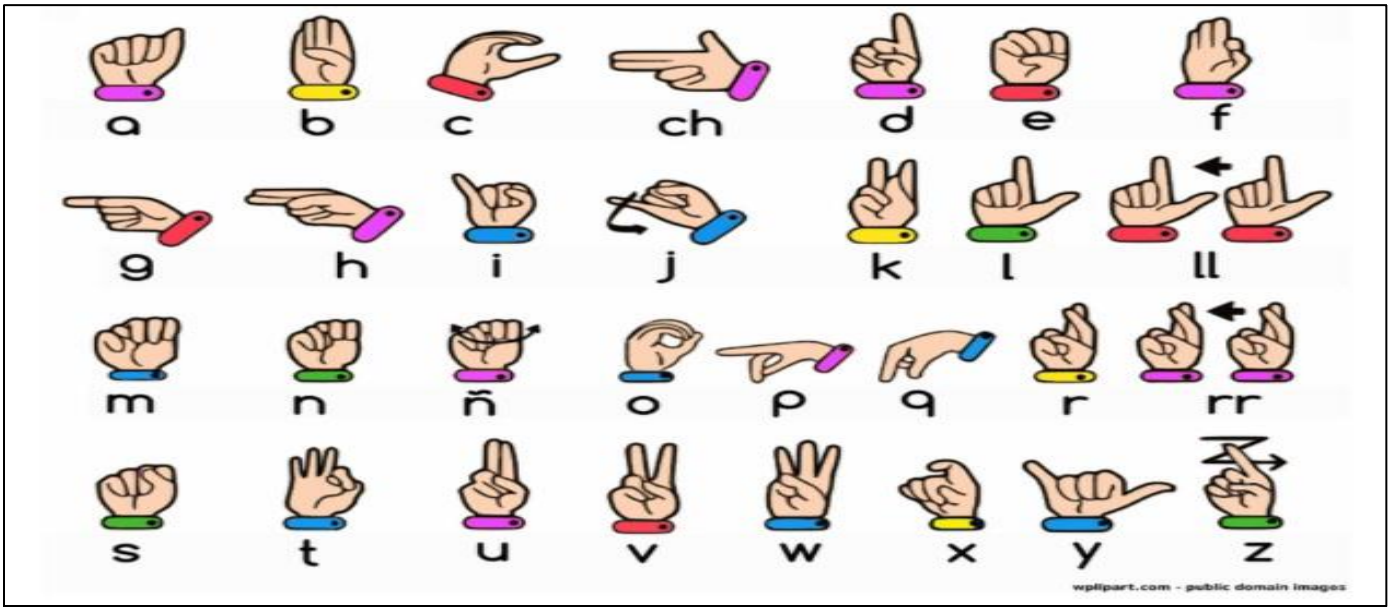

Fig 1 Alphabetic Representation using Hand Gestures

III. MOTIVATION

Although communication is a primary human function, society itself has become a least common multiple for the millions of people who have speech and hearing impairments. SIGN LANGUAGE: The deaf and hard-of-hearing community uses sign language as their primary means of communication, but it isn't widely used because most people never learn it. All of them have unmet needs, and as a result, this communication breakdown produces barriers that cause social isolation, difficulties in school, or difficulties at work. Context The problem of digital literacy has been worsened by the widening digital divide, and inclusive technologies are desperately needed to close this gap. By utilizing cutting-edge developments in artificial intelligence, machine learning, and computer vision, we hope to create a sign language translation system that does not only translate signs into text or speech but also will be able to comprehend and preserve the subtleties of each sign language and dialect while producing accurate and scalable results. Such a mechanism can facilitate participation by those with disabilities.

IV. EXISTING SYSTEM

Survey of current sign language recognition technologies and system architecture for effective real-time sign language processing Like all of them, the tiny portion of sign language translation identifies and converts sign language into text or speech. They have implemented both manual features (e.g., hand shape, motion, location, and orientation) and non-manual features (e.g., facial expressions and body postures) for accurate recognition. Commonly, deep learning models (CNN or RNN) are utilized to extract the information contained in visual data provided from camera, sensor, etc.
Some systems also incorporate wearable devices or motion capture technology to more accurately track hand movements and body posture. These devices offer rich spatial and temporal information, expanding the system's range of answered classes, and facilitating discrimination between similar signs. However, despite all of these advancements, the exiting systems still has an inability to recognize overlapping movements, handle the intricacy of dynamic signs, and scale for sign languages worldwide. Due to changes in illumination, background noise, and even unique signer styles, many systems are still restricted to controlled or laboratory settings but are applied in the real world. Recent intrinsic developments in deep neural networks and motion tracking methods have led to notable advancements in the field of recognition of sign languages.

V. FUNDAMENTAL CONCEPTS

➢ *Sign Language Recognition:*

The topic of this research is sign language recognition (SLR), the challenging problem of recognizing, understanding, and interpreting the visual gestures and expressions people use to communicate with sign language. his procedure is required since face expressions, voice direction, hand shapes, and hand movements and directions (such as up, down, left, and right) are all essential to comprehending the meaning. SLR aims to help a system recognize and convert gestures into a representation of human language, which can then be spoken or written.

➢ *Feature Extraction:*

Feature extraction is an essential stage in the process of converting the sign language into a format that can be understood by humans. When employing conventional approaches, key features are found by scanning for hand gestures using methods like SIFT and HOG. Modern techniques employ deep learning models to automatically learn the pertinent characteristics, improving accuracy and adaptability as compared to the traditional methods that relies on human created algorithms for feature extraction from images.

➢ *Classification:*

Classification begins from the extracted features of videos and identifies to which sign language symbols, words, or phrases they correspond. Old systems used a rule-based algorithm or similar statistical models; new approaches apply machine learning and neural networks for greater accuracy. CNNs are suited for static gesture recognition, and RNNs/LSTMs address sequential nature in dynamic signing.

➢ *Temporal Dynamics:*

The temporal properties of sign language are key for grasping continuous sign interpretation. This means looking at the order and timing of gestures to obtain contextual information. LSTMs, transformers, etc., are great in this regard to capture these dynamics.

➢ *Multimodal Integration:*

Multimodal systems that amalgamate the visual information with other sensory inputs, such as accelerometers or motion sensors, can enhance sign language recognition. It also adds robustness to learn and understand, particularly in the noisy or changing environments, as well as providing a richer context for the accurate interpretation.

➢ *Real-Time Processing:*

Real-time processing is necessary for practical sign language translation applications. Among the methods used to attain low-latency performance without compromising accuracy are edge computing, hardware acceleration, and model optimization.

By comprehending and decoding the aforementioned fundamental truths, our research seeks to develop a better approach that not only learns the systematization of sign language but also bridges the gap between systematized sign language and human-readable language.

VI. LITERATURE REVIEW OF THE RESEARCH PAPERS

➢ *Real-Time Sign Language Translation System*

The article, "Real-Time Sign Language Translation Systems: A review study" Maria Papatsimouli et al., is a review study focusing on real-time sign language translation. The study also pursues the idea of closing the information gap between people who are deaf/hard-of-hearing and non-deaf people through an examination of the technologies and approaches adopted in sign language recognition and translation.

- Advantage: Cost effective way for real-time conversations between deaf and non-deaf through low-cost / low-powered vision-based and sensor-based systems
- Limitation: Environmental resilience, vocabulary, and some solutions are expensive.

➢ *Research of a Sign Language Translation System Based on Deep Learning*

This survey paper by Authors: Siming He, Ridley College, St. Catharines, (Canada), We will be reviewing a paper on Deep Learning-based Sign Language Recognition for Human-Robot Interaction. The utilization of the machine learning models is done where Faster R-CNN localizes the hands, 3D CNN extracts relevant features, and an LSTM-based sequence-to-sequence model achieves recognition. By combining these technologies, they are able to validate sign language translations, completing the set with a 99% recognition rate.

- Advantage: Our model combines Faster R-CNN, 3D CNN and LSTM in an efficient manner to achieve enhanced recognition accuracy.
- Limitation: System has Limited Dataset Coverage, Not all sign language words are included in the dataset, which limits the scalability of the system.

➢ *Advancements in Sign Language Recognition*

This paper by Bashaer A. Al Abdullah, Ghada A. Amoudi and Hanan S. Alghamdi reviews the contemporary work in Sign Language Recognition (SLR) systems with special focuses on integrating Artificial Intelligence (AI) and machine learning techniques for the development of automated Sign Language Translation Systems (SLTS). The research systematically records 58 research papers that involve the deep learning techniques like Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) that have obtained a high precision in distinguishing the hand gesture.

- Advantage: Deep learning based approaches e.g. CNN, RNN have reported high accuracies for sign language recognition.
- Limitation: Limited large or diverse datasets available, especially for less common sign languages.

➢ *Sign Language Conversion to Speech with the Application of KNN Algorithm*

This paper presents the implementation of an application converting American Sign Language (ASL) gestures in real-time to both text and speech. This study is the first step towards an inclusive society where the hearing-impaired and the general population can communicate better. Convolutional Neural Networks (CNNs); K-Nearest Neighbors (KNN) After hand gestures are captures via webcam, captured gestures can either be recognized and identified or classified through KNN.

- Advantage: The application implements real-time translation where it translates sign language to text and audio for effective communication.
- Limitation: The ability to recognize an image may be affected when the image is not taken in good lighting environment.

VI. METHODOLOGY

Visual Language Interpreter (VLI) aims to plays the role of a translator between users using sign language and users who are not familiar with it. Our pipeline has a stepwise architecture through which hand gestures will neatly flow to text output and specified speech. The following stages outline the methodology adopted in this research:

➢ *Selecting a Language:*

The technology allows users to select their favourite language for the text output, making things easy for everyone. In this manner, the translated text corresponds to the user's preferred reading content. There is a simple method to choose the language before beginning gesture recognition.

➢ *Preprocessing:*

- Sound Reduction: Emits background noise and irrelevant details in the video frames, helping to streamline the gestures of the signer.
- Frame Extraction: This part grabs single frames from the video stream at a defined frame rate for smooth processing.
- Normalization: Standardizes the input data by adjusting the brightness, contrast, and resolution of the frames.
- Background Subtraction: This method removes the background to focus on the signer’s hand movements and body gestures that are useful in recognizing the sign. In this phase, the background is removed from the frames, leaving only the signer's hand and body movements visible, this is a crucial part of the process for identifying the gestures.

➢ *Classification:*

Much of this current implementation revolves around the classification module as this is the hub of the system where the learnt models on the UTD dataset are deployed to recognize sign language gestures. This module involves:
Feature Extraction: Picking out important features including hand shape, movement, orientation, and facial expressions from the preprocessed frames. Various machine learning models are put into practice, such as employing CNN and RNN networks to detect sequential dependencies by shuffling data.

➢ *Feature Extraction:*

After frames are processed, key features that characterize the gesture are defined, for example:

- Hand shape and position
- Finger orientations
- Relative motion patterns

Convolutional Neural Networks (CNNs) and computer vision algorithms like MediaPipe Hand Tracking are used for the extraction of the features, for instance, which produces a structured representation of the gesture.

➢ *Classification of Gestures:*

A trained classifier is then used to link the predicted gesture class to the appropriate sign language symbol. As a result, the model is trained using a dataset of sign language motions for accurate classification. The system is based on sign alignment, which specifies what text should be displayed for each sign.

➢ *Generating Textual and Speech Outputs:*

After the gesture is classified, it is transformed to a readable text format. Recognized text appears in real-time in the interface, as complete words and sentences determined by successive gestures. Additionally, the system may generate speech output through Text-to-Speech (TTS) synthesis, which makes it easier to communicate with people who speaks

- Algorithm: Learning Process
- Input: x, is a dimensional vector of feature
- Output: y, is the output decision

✓ Target function f : X => Y the ideal formula (Unknown)
✓ Data:($x1$,$y1$),($x2$,$y2$),.........(xn,yn) training examples
✓ Hypothesis g : X => Y formula to be used
✓ Learning algorithm $g \approx f$ final hypothesis

Volume 10, Issue 3, March– 2025 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar920

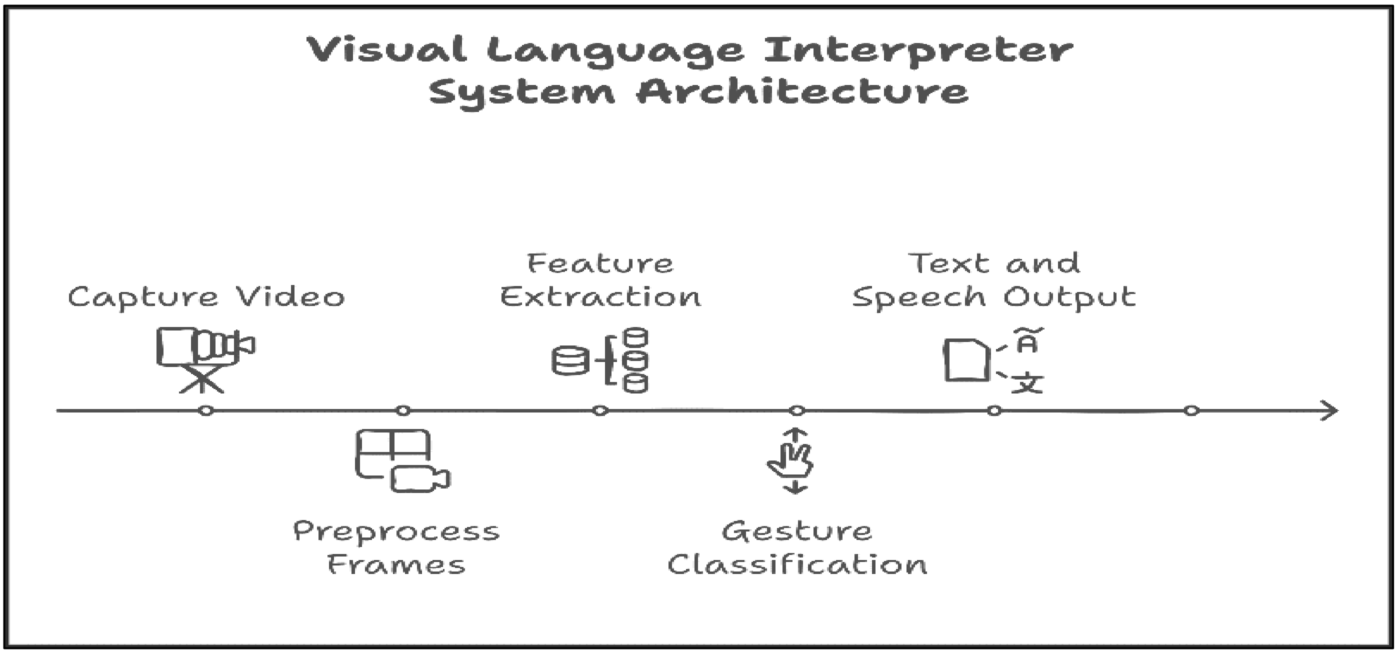

Fig 2 System Architecture of Visual Language Interpreter

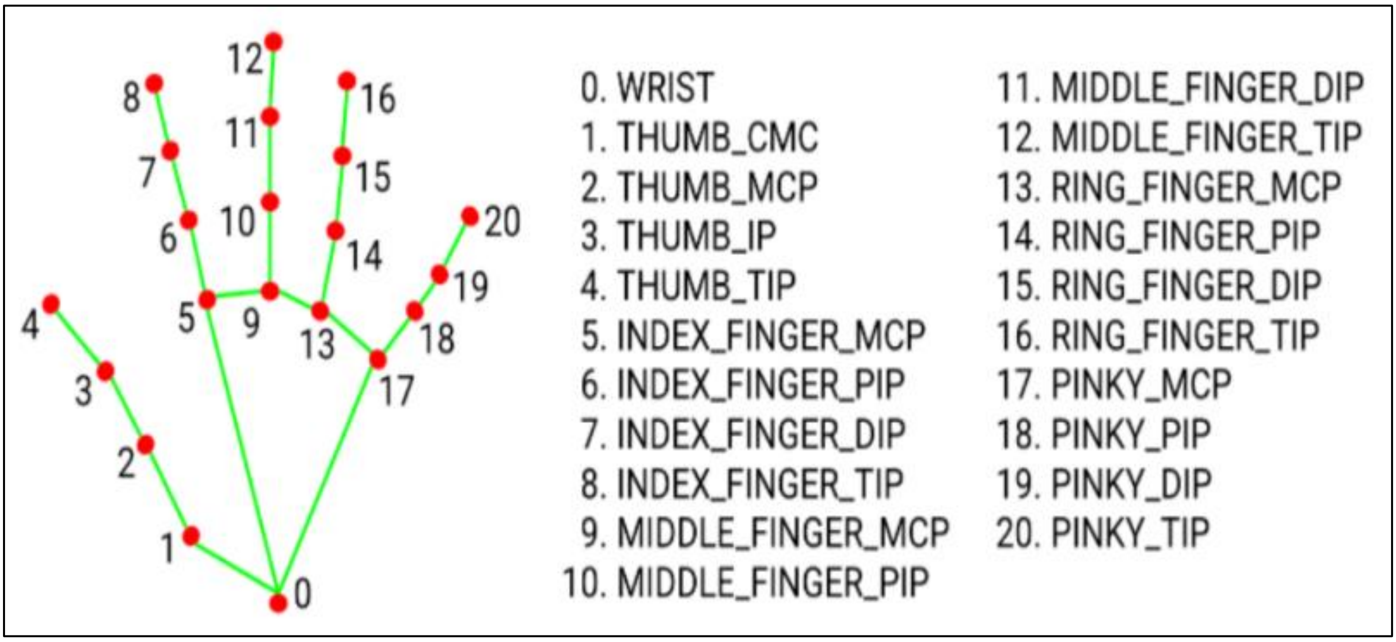

Fig 3 Media pipe Landmark System

VII. RESULT ANALYSIS

We have found a way to boost how well we can recognize sign language by changing how we classify the signs. At first, when we trained a CNN model using 26 different alphabet signs, we didn't get the results we hoped for because some hand gestures looked a lot alike. We therefore made the decision to classify these comparable indications into eight more extensive groups. This made it simpler to grasp and reduced misunderstandings among each group. We construct a probability distribution for each group, and our forecast sign is the sign with the highest likelihood. We also do some math on the hand landmarks, which allows us to tell the signs apart more accurately within those groups. This step-by-step process really boosts our recognition accuracy.

Following extensive testing, we discovered that our model achieves an astounding 97% accuracy in a variety of background and lighting conditions. Accuracy can reach 99% under ideal circumstances, such as clean backgrounds and bright lighting. This demonstrates the strength and dependability of our approach to real-time sign language interpretation.

- Insight: When compared to other approaches, the fingerprint system exhibits the best balance between precision and recall, as seen by its highest F1 Score of 99.24%.

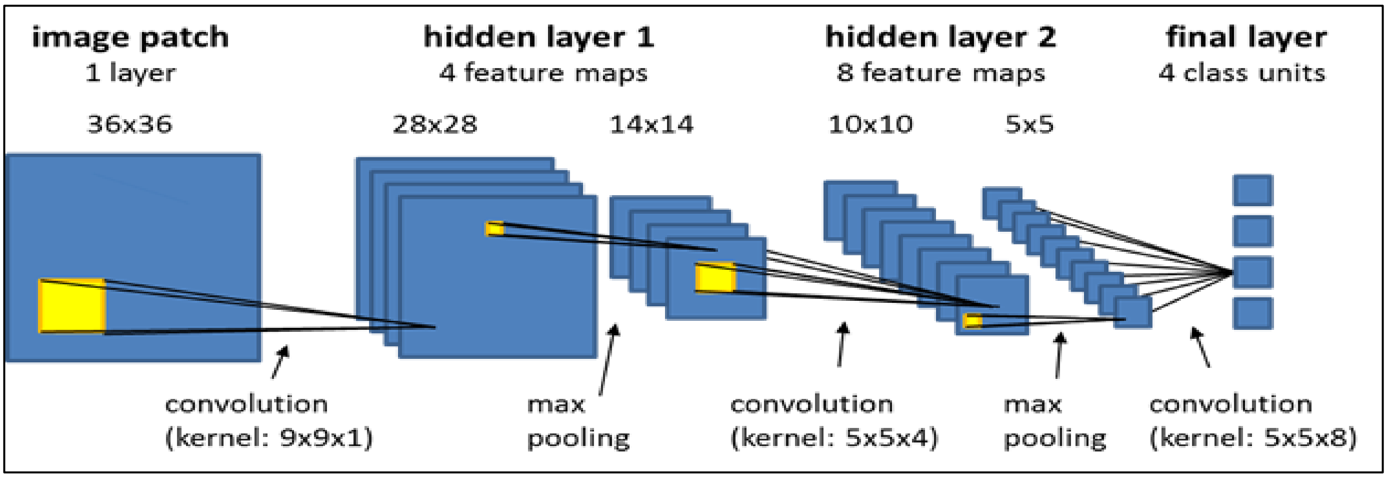

Fig 4 Convolution layer

A little window size (usually 5*5) that reaches the depth of the input matrix is present in the convolution layer. As I proceed, I will produce a 2-Dimensional activation matrix that provides the matrix's response at each geographic location. Insight: The fingerprint system shows the highest success rate, indicating superior reliability and accuracy.

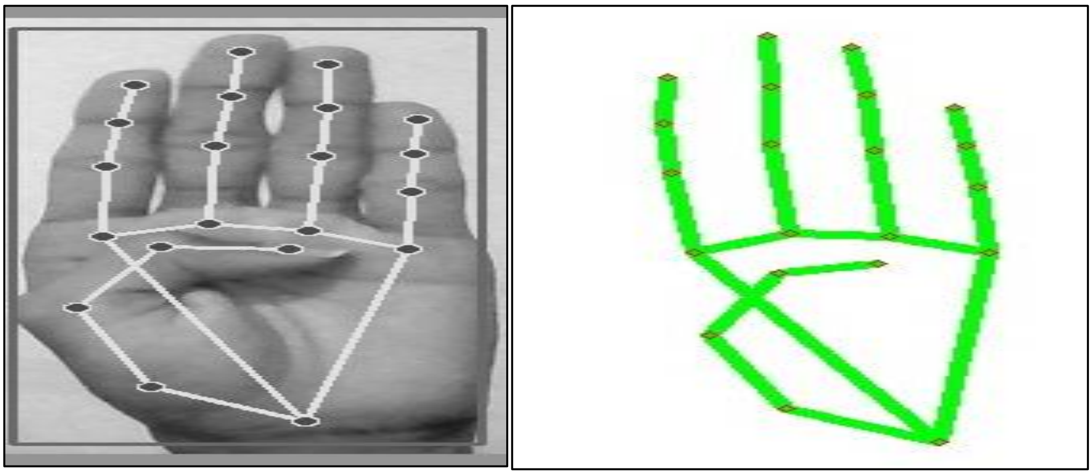

Fig 5 20-point palm detection

The mediapipe library and OpenCV had been a major help for obtaining these landmark points, and they were subsequently displayed on a plain white background. By doing this we tackled the situation of background and lighting conditions because the mediapipe library gives us the landmark points in any background and mostly in any lighting conditions

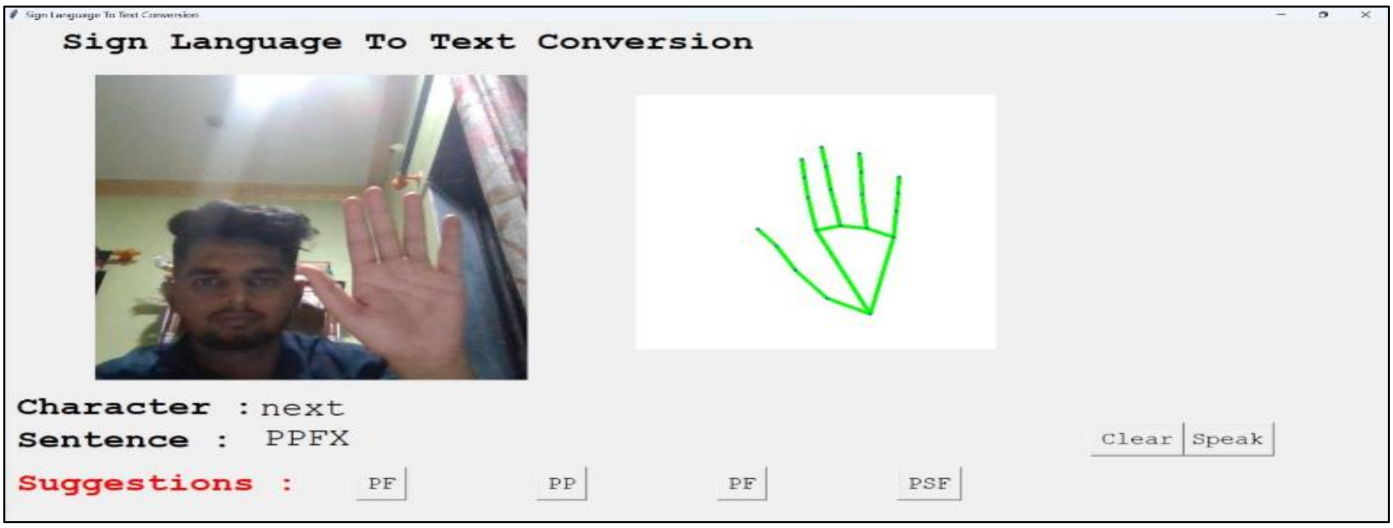

Fig 6 Real-Time Sign Language Recognition Interface

VIII. CONCLUSION

Convolutional Neural Networks (CNNs), a potent deep learning technique, allowed us to create a system for real-time sign language detection and translation. All this while bringing both the hearing-impaired community and those who depend on spoken language closer together through ease of communication and reducing social isolation. Combining modern technologies such as utilizing MediaPipe for tracking hands/landmarks, and are able to maintain a recognition rate higher than 97% across multiple environmental settings (illuminations, background noise, etc.). However, the ability of the system to preprocess video frames, extract key features, and classify gestures into contextual text and speech outputs shows promise for real-world applications. The method of classifying comparable motions into more general categories significantly improves identification accuracy by removing ambiguity and bolstering the system's resilience. The findings indicate that a second round of sequence ranking was conducted after a 98.7% accuracy rate was attained in a range of settings.

REFERENCES

[1]. Maria Papatsimouli, Konstantinos-Filippos Kollias, Lazaros Lazaridis, George Maraslidis, Herakles Michailidis, Panagiotis Sarigiannidis and George F. Fragulis,”Real-Time sign language Translation system” 2022 11th International Conference on Modern Circuits and Systems Technologies (MOCAST) | 978-1-6654-6717-9/22/$31.00©2022IEEE|DOI:10.1109/MOCAST548 14.2022.9837666

[2]. BASHAER A. AL ABDULLAH, GHADA A. AMOUDI, AND HANAN S. ALGHAMDI, "Advancements in Sign Language Recognition: A comprehensive review and future prospects." IEEE Access, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3457692.

[3]. Y. Cheng, W. Shang, L. Zhu and D. Zhang, "Design and implementation of ATM alarm data analysis system," 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), Okayama, Japan, 2023, pp. 1-3.

[4]. Rajanishree M, Nadeem Ahmed N, Yashvi Panchani, Shreyaa Aravindan1 and Viraj Jadhav Department of Computer Science and Engineering, School of Engineering and Technology, Bengaluru, Karnataka, India.,” Sign Language Conversion to Speech with the Application of KNN Algorithm” 2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) | 978-1-6654-6941-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/I-SMAC55078.2022.998742

[5]. Siming He Ridley College, St. Catharines, Canada, “Research of a Sign Language Translation System Based on Deep Learning” 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM). | 978-1-7281-4691-1/19/$31.00 ©2019 IEEE DOI 10.1109/AIAM48774.2019.00083.

[6]. Gartner, C. (2021). "The Future of Sign Language Recognition: Bridging the Communication Gap." Technology Review.

[7]. Kumar, P., Singh, A., & Mahato, H. (2020). "A Comprehensive Study on Sign Language Recognition Systems." Journal of King Saud University - Computer and Information Sciences.

[8]. J.M.Power, G. W. Grimm, and J.-M. List. "Evolutionary dynamics in the dispersal of sign languages," Roy. Soc. Open Sci., vol. 7, no. 1, Jan. 2020, Art. no. 191100.

[9]. J. Rowe, "Artificial intelligence and the future of sign language." LinkedIn, 2017. Accessed: Nov. 3, 2021.

[10]. U. Farooq, M. S. M. Rahim, N. Sabir, A. Hussain, and A. Abid, "Advances in machine translation for sign language: Approaches, limitations, and challenges." Neural Computer Application, vol. 33, no. 21, pp. 14357-14399, Nov. 2021.

[11]. M.A. Abdel-Fattah, "Arabic sign language perspective." 1. Deaf Snud. Deaf Educ. vol. 10, no. 2.pp. 212-221. Apr. 2005.

[12]. B. S. Parton, "Sign language recognition and translation: A multidisci. plined approach from the field of artificial intelligence," J. Deaf Stud. Deaf Educ. vol.11. no. 1. pp. 94-101, Oct. 2005.