SIGN LANGUAGE DETECTION AND TRANSLATION APPLICATION

Athul Krishna
Dept. of Computer Science and Engineering
Sahrdaya College of Engineering and Technology
Thrissur, Kerala
athulkrishna2k3@gmail.com

Ashwin Joy
Dept. of Computer Science and Engineering
Sahrdaya College of Engineering and Technology
Thrissur, Kerala
ashwinjoy012@gmail.com

Indralal Shaiju
Dept. of Computer Science and Engineering
Sahrdaya College of Engineering and Technology
Thrissur, Kerala
indralalshaiju@gmail.com

Sweeto Paul
Dept. of Computer Science and Engineering
Sahrdaya College of Engineering and Technology
Thrissur, Kerala
sweetopaul99@gmail.com

Krishnadas J
Dept. of Computer Science and Engineering
Sahrdaya College of Engineering and Technology
Thrissur, Kerala
krishnadasj@sahrdaya.ac.in

Anly Antony
Dept. of Computer Science and Engineering
Sahrdaya College of Engineering and Technology
Thrissur, Kerala
anlyantony@sahrdaya.ac.in

***Abstract*—This paper introduces an innovative, user-centric system for real-time sign language detection, integrating advanced hand-tracking technology with AI-driven algorithms to facilitate smooth communication between all the individuals using sign language and those who do not. The system translates sign language into English or other languages in real-time, offering broad language support to make it versatile in global contexts. Designed for high accuracy and adaptability, the system leverages real-time hand movement tracking and is trained on large-scale, diverse datasets to handle various gestures, lighting conditions, and environments effectively. Customizable features allow users to tailor the system to their needs, ensuring that only one participant needs to learn the sign language for effective communication. By providing a scalable, robust solution, this project aims to enhance communication accessibility and bridge language barriers for diverse users and communities.**

***Index Terms*—Sign Language Detection, Real-time Translation, Hand Tracking Technology, AI-Driven Algorithms, Gesture Recognition, Accessibility in Communication, Multilingual Support, Deep Learning, Natural Language Processing (NLP), Machine Learning (ML), Inclusive Technology, Computer Vision, Human-Computer Interaction (HCI), Dynamic Sign Recognition**

I. INTRODUCTION

In recent years, the involvement of artificial intelligence (AI) and computer-based vision technologies has led to groundbreaking advancements in human-computer interaction. One area where these advancements have had a profound impact is in hand tracking and gesture recognition systems. These systems have enabled intuitive and immersive interactions with digital environments, ranging from virtual reality experiences to augmented reality applications and beyond. In this paper, we introduce an AI-based hand tracking sign translation system utilizing hand gestures. This innovative system leverages the power of deep learning algorithms to accurately track hand movements and interpret them into meaningful translations. By harnessing the capabilities of AI, we aim to bridge the gap between human gestures and digital interactions, opening up new possibilities for accessibility and communication. Conventional translation methods that depend on human interpretation or pre-made databases are frequently ineffective and unable to take into account sign language's dynamic character. The use of cutting-edge technologies capable of effectively recognizing, processing, and translating hand gestures and facial expressions is necessary to meet the increasing need for real-time, automatic translation solutions. Advanced techniques in computer vision, machine learning (ML), and natural language processing (NLP) are being investigated to create more accurate and efficient sign language translation systems in order to overcome these difficulties. The system can track hand movements, identify gestures in real time, and convert them into speech or text thanks to these technologies.

II. PROBLEM DEFINITION

This paper develops a real-time sign language recognition system which, because of the application of artificial intelligence and state-of-the-art hand-tracking technologies, will be in a position to give perfect, efficient translations. It will allow recognition of massive amounts of movements within sign languages, with the minimum latency as possible, having good responsiveness as well as good adaptability on changes in the signing styles or the user's preferences. It has all guarantees towards inclusion and accessibility to very broad customers. The initiative brings sign language recognition to the forefront by incorporating modern technology with accessibility, diversity, and user experience. This project will enhance meaningful communication and improve societal inclusion through user-centric designs and High real-time performance.

III. Motivation

This study responds to the urgent need to bridge the communication gap between deaf and hard of hearing (DHL) people and those who do not use sign language. Sign language has a peculiar structure that makes it difficult to communicate effectively with people who use spoken or written language. But at the same time, vision-based communication, for instance, in the form of sign language recognition that converts the signs into real-time text, might bridge the gap by making both understand each other.

Important reasons are:

- Improved Access to Communication: The ability to create an immediate conversion from sign language into text would ensure Deaf and mute persons, along with the hearing population, easily connect more efficiently and cross over the barriers that hinder the line of communication.

-Inclusive technology promotes accessibility and inclusivity as it enables interaction without having to learn each other's language with an intuitive HCI that can be understood by various sign languages.

-Global Language Support: Various sign languages, including American Sign Language, British Sign Language, Indian Sign Language, French Sign Language, and Japanese Sign Language, are spoken in different regions worldwide.. Our project will support more than one sign language to adapt the system to reach the whole world.

-Improved User Engagement: The system can provide live translation that responds to different sign languages and the preferences of the users using computer vision and gesture recognition technology. This makes conversation meaningful.

-The project aims to contribute to a more inclusive society by ensuring effective and barrier-free communication between people of diverse linguistic origins.

IV. Objectives

The objectives of this work are:

- Create applications where there is exact identification and tracking of hand and finger movement in real time using reliable hand-tracking algorithms.
- Use AI-based technology to identify and convert gestures in sign language into either spoken words or commands on a game controller.
- With such building of the system to cater for users' preference and their individual signing techniques, we ensure inclusivity and access.
- Involving the Deaf community also serves as a means to encourage the use of this technology and bring out awareness as it may encourage people to bring together inclusiveness through breaking communication barriers.
- Multiple sign languages should be supported to honor and take into account regional and cultural differences in signing styles.

V. LITERATURE SURVEY

[1] R Kumar et al. compares and contrasts several methods and algorithms for identifying sign language. It examines and assesses the advantages and disadvantages of conventional methods such as deep learning (DL) and machine learning (ML). This paper mainly discloses the potential problems with precision and efficiency associated with real applications of these algorithms and discusses major challenges in handling variant hand shapes along with dynamic styles of signing. Above all, the work under discussion focuses on explaining the importance of context for recognition of sign language motions and makes a summary to explain the procedures which enhance performance by increasing gestures' segmentation or recognition precision.

[2]S Dalal et al. contrasts a variety of methods used for sign language recognition. They vary from wearable sensor-based methods to vision-based methods. It has compared the advantages and disadvantages of each of them. I has also emphasized on their efficiency in real-time translation of sign languages. Besides the deep learning architectures CNN and RNN, it considers techniques in 2D and 3D image processing. The paper concludes by highlighting the need for hybrid systems that combine multiple strategies to address various signing situations and environments.

[3]M S Amin et al. proposed a range of sensor technologies, such as cameras, depth sensors,different wearable technology like gyroscopes and accelerometers, that are utilized for sign language recognition. The study examines the ways in which various sensors enhance the precision and resilience of sign language translation systems.. It describes the difficulties in choosing the ideal sensor for an environment and contrasts passive sensors such as RGB cameras with active sensors like depth cameras (such as Kinect). The paper further elaborates about different sensor fusion techniques, which gather information from multiple sensors to enhance gesture recognition and minimize errors due to occlusions or other extrinsic factors.

[4]M R Cassim et al.Specializes in creating an accessible and cost-effective sign language to voice translator that is portable.The authors talk about how at the current price of resources its been difficult to create reasonably priced systems without sacrificing usability or accuracy. The cheap sensors along with computer vision algorithms transform hand signs into text. A following TTS module delivers the inputted text into speech. Given that this project emphasizes portability and

processing in real-time, this is a proper device for daily use in many different scenarios.

[5]A Makkar et al. presents an open-source platform for translating ASL. The study offers a novel method for classifying sign language gestures using time-series data by utilizing deep learning models that process continuous motion patterns in real-time. The study emphasizes how well Long Short-Term Memory (LSTM) networks and Recurrent Neural Networks (RNNs) capture temporal relationships in sign language gestures. The project has an open-source design hence, it fosters teamwork and makes it a useful resource for future studies on real-time sign language translation.

[6]O Koller et al. Provides a comprehensive examination of the latest advancements in sign language recognition. It also talks about both conventional approaches and the recent developments. It examines different strategies, such as sensor-based techniques, optical character recognition (OCR), and other deep learning-based models. They mentions the application of these techniques in practical situations and performs research on their effectiveness across various datasets. There are a wide range of hand shapes, signing techniques, and other factors that impact recognition accuracy so the research discusses the difficulties in training models on varied datasets.

[7]V S Attili et al. uses computer vision and deep learning methodologies to perform real-time sign language translation. The paper is emphasized on the use of RNNs for temporal sequence recognition along with CNNs for image recognition and hand gesture detection. The system is designed to translate in real time so that it reduces latency and performs smooth communication. The article also discusses the challenges of real-time processing, computing complexity, and maintaining high accuracy in a variety of environmental circumstances.

[8] M. Papatsimouli et al. investigates how Internet of Things (IoT) technology can be integrated with sign language translation systems. The study explores how IoT devices might improve sign language recognition systems by facilitating multi-sensory inputs, real-time feedback, and remote monitoring. Some examples of IoT devices that can be used with wearables and smart cameras to enhance the gesture recognition and, thus, expand the capabilities of the system beyond the typical computer settings include wearables and smart cameras. This research demonstrates how the possibility of creating more sophisticated and versatile sign language interpreting applications that are sensitive to a number of conditions can be realized.

[9]J. Li et al. states the challenges that render sign language translation systems ineffective and provides suggestions on how to improve the effectiveness of the systems with focus on enhancing the accuracy and usefulness of the systems. The paper considers various advancements in machine learning algorithms through which computers are able to adapt to different languages and forms of signing such as the transfer learning. The article also describes how user feedback can be incorporated into the system to enhance its performance, and thus how the sign language translation can remain versatile enough to suit different users and conditions.

[10]M Fernando et al. depicts that Hu moments, or rather the 7 Images invariant moments, are suitable for real time image processing applications such as sign language translation. The paper devises a new way of recognizing hand gestures through the combination of the Hu moments and traditional image processing techniques. This allows for a quick and accurate way of interpreting hand signs, and it is expected to be cost-effective as well as suitable for real time applications. To enhance efficiency of gesture recognition systems, the authors explain the importance of feature extraction and image preprocessing.

VI. Methodology

The methodology outlines the systematic process of developing a robust hand gesture recognition system. It integrates State-of-the-art machine learning technology for feature extraction, data preprocessing, model training, and evaluation to ensure accurate classification of hand gestures.

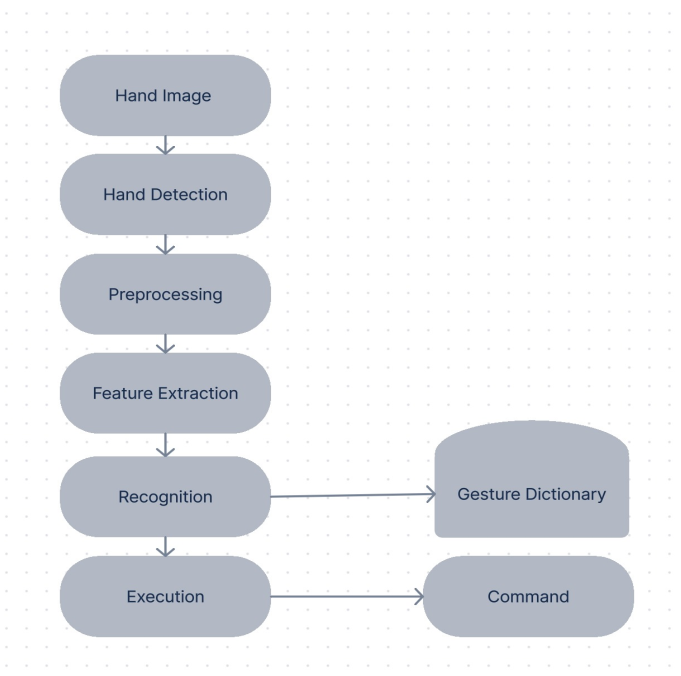

Fig. 1. flowchart

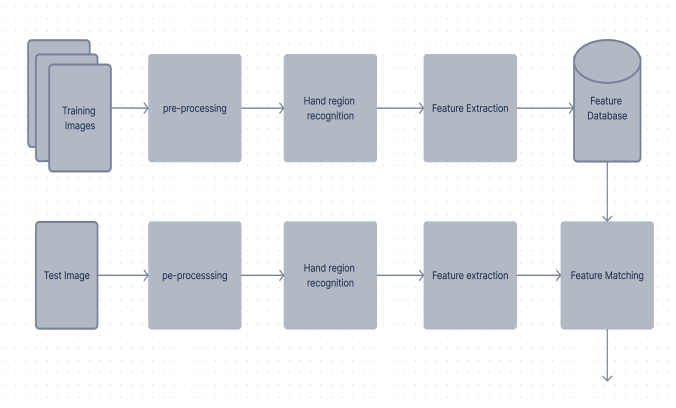

Fig. 2. Architecture Diagram

A. Data Collection and Preprocessing

1) Data Collection:

- **Video Capture:** Gesture videos were recorded using OpenCV with a standard webcam setup.
- **Dataset Organization:** Videos are stored in a directory structure with subfolders (`0`, `1`, ..., up to `number_of_actions - 1`) representing distinct gestures. Each subfolder contains multiple video sequences corresponding to the respective action.

2) Data Labeling:

- Each subdirectory is assigned a label corresponding to its action class. For example:
  - Subdirectory `/0` represents action `0` (e.g., ”Wave”).
  - Subdirectory `/1` represents action `1` (e.g., ”Thumbs Up”).

3) Preprocessing Steps:

- **Feature Extraction:**
  - Hand landmarks are detected for each frame in a video using Mediapipe.
  - The medipipe hand model detects 21 landmarks for each hands.
  - Each hand will contribute 21 landmarks,also it has two cooedinates x and y therefore 21*2=42 features per hand.Therefore if two hands are detected in a frame total features extracted will be 84 featurers.
  - Thus in case of one hand detection,42 features are extracted
- **Data Cleaning:** Removal of incomplete or corrupted videos to ensure data quality.
- **Feature Scaling:** StandardScaler is applied to normalize features, making the model less sensitive to outliers and improving convergence during training.
- **Dataset Partitioning:**
  - **Training Set:** 80% of the available data is utilized for training the model.
  - **Testing Set:** 20% of the data populated is used for evaluating model performance.

B. Machine Learning Model Training

1) Feature Representation: Each gesture video is represented as a sequence of feature vectors containing normalized coordinates of hand landmarks. These vectors capture the spatial configuration of the hand across frames.

2) Training Process:

- **Algorithm Selection**
  - SVM works well with smaller datasets as it provides efficient results even with limited data.
  - The code uses a linear SVM, meaning it attempts to find a linear decision boundary that separates different classes. Linear SVMs are simpler to train and computationally more efficient.
  - If the data is not linearly separable, SVM can be modified to use non-linear kernels like RBF, polynomial, etc.
  - SVMs are simple to understand and interpret compared to more complex models like deep neural networks.
  -
- **Algorithm:** A Support Vector Classifier (SVC) with a linear kernel is selected for its effectiveness in handling high-dimensional feature spaces. Probability estimation is enabled to allow confidence-based predictions.
- **Hyperparameter Tuning:** Grid search or random search techniques may be used to optimize model hyperparameters like `C` (regularization) and kernel type for improved accuracy.

C. Dataset Considerations and Augmentation

- **Dataset Size:** The dataset should contain sufficient samples for each class to ensure effective learning. Ideally, 50–100 videos per class are recommended.
- **Data Augmentation:** To increase data diversity and robustness, techniques such as flipping, rotation, and adding noise to videos can be applied.
- **Class Imbalance:** To address class imbalance, oversampling (e.g., SMOTE) or undersampling techniques may be used, ensuring equitable representation of all classes.

TABLE I
DATASET BREAKDOWN BEFORE AND AFTER AUGMENTATION

Class (Sign)	Videos per Class (Before Augmentation)	Augmentations per Video	Total Videos (After Augmentation)
Sign 1	10	5	50
Sign 2	10	5	50
Sign 3	10	5	50
Sign 4	10	5	50
Sign 5	10	5	50
Sign 6	10	5	50
Sign 7	10	5	50
Sign 8	10	5	50
Total	**80**	**5**	**400**

D. Summary of Workflow

1) Capture gesture videos and organize them into labeled subdirectories.
2) Extract normalized features using Mediapipe hand landmark detection.
3) Scale and split data into training and testing sets.
4) Train an SVC model to classify gestures based on feature vectors.
5) Evaluate the model's performance using accuracy, precision, recall, and other metrics.

This methodology ensures a systematic approach to developing a high-performing gesture recognition system.

VII. EXPERIMENT AND RESULT ANALYSIS

A. Evaluation of Performance

With the integration of WebSocket technology, the application enables stable communication for real-time data exchange between the client and server, processing frames seamlessly. This design minimizes latency, ensuring consistent and responsive gesture detection during video streaming.

To optimize resource management, the application skips frames effectively and uses threading for TTS responses, which greatly reduces processing lag and results in a smooth and responsive user experience.

For two reasons, media pipe applications work best in appropriate lighting conditions: first, media pipe has significant issues functioning correctly in low lighting conditions which might lead to misclassifications, and second, the application detection or recognition of hand gestures may not work. Moreover, excessive talking using different languages may lead to high memory consumption, application caching and selective text to speech application may improve efficiency and speed.

This application takes a sophisticated, user-centered approach to hand gesture recognition that will give accurate results, allows for multiple languages and is non intrusive for end users, and this makes it suitable for interactive applications.

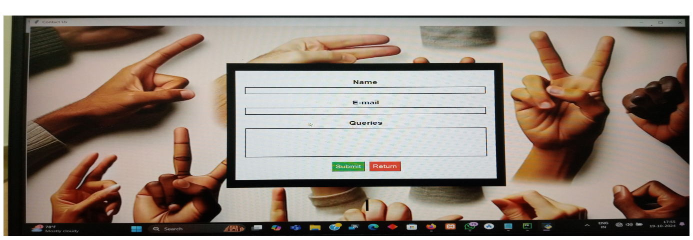

Fig. 3. Login Page

Fig. 4. Sign Detection

Fig. 5. Sign Detection

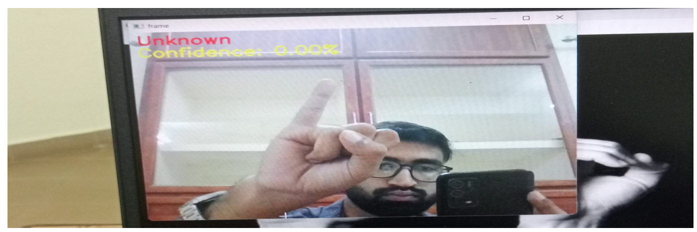

Fig. 6. Invalid Sign Detection

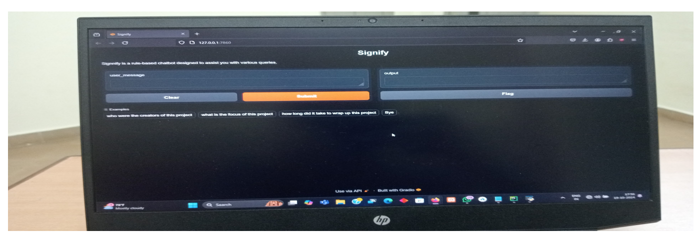

Fig. 7. Chatbot

VIII. CONCLUSION

In summary, this hand gesture recognition app demonstrates how well computer vision and natural language processing can be used together so that an approacable and real-time experience is created. The application utilizes MediaPipe for hand tracking, which is associated with a TensorFlow model for categorizing gestures in order to make it highly precise in gesture recognition with low latency, thus making it responsive and applicable to various purposes. Multilingual TTS enables fluent communication among users from different linguistic backgrounds, and hence it's very inclusive. Combining WebSocket communication with Tkinter GUI design creates a dependable and intuitive interface that easily accommodates real-world usage situations. Even though there are certain drawbacks, like such reliance on illumination and possible memory consumption issues with constant TTS answers, these can be minimized with additional optimization approaches, such as the use of adaptive lighting and effective memory management procedures. There is definitely much room to improve this application in the future. More various sign language motions can be added in the gesture vocabulary, and also, the adjustment capacity of this model towards diverse users' signing styles can be enhanced, so elements like emotion recognition can also be added in order to be able to achieve better communication. Additionally, expanding compatibility across more platforms and gadgets, such as wearables and mobile apps, may improve usability and accessibility. Overall, this project lays a solid basis for interactive applications that use gesture-based controls, paving the way for further advancements in accessibility and control systems that rely on gestures.

REFERENCES

[1] R. Kumar, A. Sinha, A. Bajpai, and S. Singh, "A comparative analysis of techniques and algorithms for recognising sign language," *arXiv preprint arXiv:2305.13941*, 2023.

[2] S. Dalal, R. Kacheria, and V. Venkataramanan, "A comparative study on sign language recognition methods," in *2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*. IEEE, 2022, pp. 1–7.

[3] M. S. Amin, S. T. H. Rizvi, and M. M. Hossain, "A comparative review on applications of different sensors for sign language recognition," *Journal of Imaging*, vol. 8, no. 4, p. 98, 2022.

[4] M. R. Cassim, J. Parry, A. Pantanowitz, and D. M. Rubin, "Design and construction of a cost-effective, portable sign language to speech translator," *Informatics in Medicine Unlocked*, vol. 30, p. 100927, 2022.

[5] A. Makkar, D. Makkar, A. Patel, and L. Hebert, "Signspeak: Open-source time series classification for asl translation," *arXiv preprint arXiv:2407.12020*, 2024.

[6] O. Koller, "Quantitative survey of the state of the art in sign language recognition," *arXiv preprint arXiv:2008.09918*, 2020.

[7] V. S. Attili, A. Kottapalli, A. Aithal, A. K. Pawa, and A. M. Joshi, "Real-time sign language translation using computer vision and machine learning," in *2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC-ROBINS)*. IEEE, 2024, pp. 703–709.

[8] M. Papatsimouli, P. Sarigiannidis, and G. F. Fragulis, "A survey of advancements in real-time sign language translators: integration with iot technology," *Technologies*, vol. 11, no. 4, p. 83, 2023.

[9] J. Li, J. Xu, Y. Liu, W. Xu, and Z. Li, "Enhancing the applicability of sign language translation," *IEEE Transactions on Mobile Computing*, 2024.

[10] M. Fernando and J. Wijayanayake, "Novel approach to use hu moments with image processing techniques for real time sign language communication," *arXiv preprint arXiv:2007.09859*, 2020.