Real-Time Sign Language Detection and Recognition Using Live Camera Interface with Multilingual Translation and Speech Synthesis

Palaash Surana
Department of Computing Technologies SRM Institute of Science and Technology, Kattankulathur, Chennai,603203, Tamil Nadu, India pa0770@srmist.edu.in

Kamya Ojha
Department of Computing Technologies SRM Institute of Science and Technology, Kattankulathur, Chennai,603203, Tamil Nadu, India ka4242@srmist.edu.in

Anita R
Department of Computing Technologies, Faculty of Engineering and Technology, School of Computing SRM Institute of Science and Technology, Kattankulathur, Chennai,603203, Tamil Nadu, India anitar@srmist.edu.in

Abstract— Technology for sign language recognition greatly helps deaf and hearing populations to communicate. We propose in this study a computer vision-based real-time sign language recognition and translating system. The system receives geographical and temporal data from hand, face, and stance markers using MediaPipe for real-time gesture detecting. After that, a Random Forest classifier handles these characteristics not just for motion labelling but also for text translation. English, German, Hindi, Japanese, Tamil, and Telugu are only a few of the several languages the translated work finds use in. Then, to guarantee flawless communication, text to speech (TTS) technology synthesizes the translated text into audio. Unlike current methods relying just on image processing and suffering from latency, our system blends advanced algorithms, high-quality datasets, and enhanced sensors to improve accuracy and response speed. Designed for real-time applications, the suggested solution is fit for assistive devices, educational tools, daily communication, therefore encouraging accessibility and inclusion for people with hearing and speech problems. The proposed model is implemented on Python. Then the efficiency of the proposed approach is compared with existing models. The accuracy, precision, recall and f1 score of the suggested approaches are all assessed. The proposed Real-Time SLDR- with Multilingual Translation and Speech Synthesis attains 0.94 accuracy, 0.94 F1-score, 0.94 recall, 0.95 precision.

Keywords— Sign Detection, Random Forest Classifier, translation, Speech, text to speech.

I. INTRODUCTION

Communication gaps between the deaf and hearing populations cause great challenges that occasionally lead to social exclusion and limited opportunities for people with hearing and speech problems. Especially used among the groups of deaf and mute people, sign language is a disciplined visual-gestural communication tool. Still, a big challenge exists: the general public is not familiar with sign language, which emphasizes how urgently automated translating technologies that can properly close this communication gap are needed.

This work presents real-time sign language recognition and translational system based on computer vision and machine learning. The system compiles gestures by extracting spatial and temporal data from camera video input using hand, facial, and posture markers. MediaPipe guarantees accurate and efficient gesture classification by means of an ensemble of decision trees handled by a Random Forest classifier, so attaining real-time gesture recognition. Unlike existing systems depending just on image processing and have great latency, this method combines advanced feature extracting techniques, great datasets, and optimal classification algorithms to improve recognition accuracy and response speed.

Among its several contributions is the ability of this system to translate recognized sign language gestures into several languages, including English, German, Hindi, Japanese, Tamil, and Telugu. Originally written, the observed gestures are first converted into the selected language before being combined using text to speech (TTS) technology into audio. This multilingual translating function ensures that the system is adaptable enough for several language needs, hence it is a helpful tool for world communication.

Real-time applications are suited for the Random Forest model since it can effectively control complicated feature interactions, lower overfitting, and offer fast inference times all of which help this framework. Designed for daily contacts, assistive technologies, and instructional tools, this approach supports accessibility and inclusiveness for persons with speech and hearing problems. This approach closes the communication gap so enabling a society more inclusive and linked where language is no more of a barrier to contact.

II. RELATED WORK

This section does a survey of the literature on the relevant work. The first perspective examines the literature on recognition and translation of sign languages as the proposed work focuses on creating a machine translation system for sign language datasets. The second viewpoint deals with translating video image into sign languages. Numerous fundamental methods of machine translation have been developed. Each ofthese viewpoints offer insightful information about current approaches and difficulties, which aids in placing the suggested work in the context of translation research as a whole. By looking at these areas, we hope to draw attention tothe particular shortcomings and technological developments that our system solves, especially when it comes to sign language translation and recognition in a variety of language contexts.

In [1], Núñez-Marcos (2023) investigated machine learning techniques for sign language machine translation in order to raise translating accuracy. Published in Expert Systems with Applications (Vol. 213), their work obtained a BLEU score of

46, therefore demonstrating the effectiveness of their approach in sign language translation.

In [2], Lianyu Hu (2024) presented a scalable frame resolution technique for effective continuous sign language detection. Published in Pattern Recognition (Vol. 145), their work applied CNNs to dynamically switch among several resolutions, hence enhancing network efficiency.

From an interdisciplinary standpoint, Bragg (2019) investigated sign language recognizing, generating, and translating [3]. Published in ASSETS, their CNN approach work generated a Word Error Rate (WER) of 39.6%.

For deaf-mute people in [4] Suharjito and Anderson (2017) created a sign language recognition application system. Presented at ICCSCI, their work used CNN, SVM, and HMM approaches to attain respective recognition accuracies of 91%, 97%, and 86%.

In [5] Rao (2023) presented a sign language recognition system leveraging MediaPipe and LSTM. Published in ICICCS, their study showed an accuracy of 85–90%, therefore underlining the success of LSTM-based models for gesture recognition.

For sign language recognition and translation in [6], Gan and Jiang (2023) presented a contrastive learning method. Published in IJCAI, their work obtained a BLEU score of 21.79% by using CSLR and SLT models.

For sign language transliteration in [7], Angelova (2022) used neural machine translation techniques Published in the Student Research Workshop, their work used RNN models and earned a BLEU score of 20.

Camgoz (2018) proposed in [8] a neural sign language translation model using LSTM-based models. Showing their work at an IEEE conference, their BLEU result came out to be 43.

Chen (2022) created in [9] a multi-modality transfer learning method for sign language translation. Published at the IEEE/CVF conference, their work combined ML, CNN, and mBART methods to get a BLEU score of 44.

Rastgoo (2021) reviewed sign language production in [10]. Presenting their work at the IEEE/CVF conference, their investigation on ML and CNN-based methods revealed a BLEU score of 52, therefore emphasizing the value of multi-modality learning.

Ajay S. (2021) built an Indian Sign Language detection system in [11] using an Arduino Random Forest classifier. Published in IEEE (CONECCT), their study demonstrates Random Forest's feasibility for sign language detection using a BLEU score of 33.

In [12], Sunanda Das (2022) presented a hybrid technique combining CNN and a Random Forest classifier for Bangla Sign Language identification. Published in Elsevier, their study underlined the accuracy, precision, recall, and F1-score values of 91.67%, 93.64%, 91.67%, and 91.47%, respectively, of the combined method with accuracy.

Using an input-process-based approach, Suharjito and Ricky Anderson (2017) examined Sign Language Recognition Application Systems for Deaf-Mute People in [13]. Published at the 2nd International Conference on Computer Science and Computational Intelligence (ICCSCI 2017) in Bali, Indonesia, their work assessed several classifiers with 91% for CNN, 97% for SVM, and 86% for HMM reporting accuracy rates.

III. PROPOSED WORK

The system that has been proposed is a real-time framework for the recognition and translation of sign languages. It is designed to ensure that the processing is effective and low-latency by incorporating computer vision, machine learning, and multithreading. The system facilitates seamless communication between the deaf and hearing communities by capturing video input, recognizing sign language gestures, translating them into multiple spoken languages, and generating audio output. In Fig 1.1 shows all the process occurring in the model and Fig 1.2 shows the user interface.

3.1. Processing of Input

Through a camera, the device captures real-time sign language gestures, therefore guaranteeing a flawless engagement. Essential for the exact recognition of motions, spatial and temporal aspects from important markers—hands, its finger position, distance and angle are extracted using MediaPipe. These tools enable the efficient understanding of sign language by helping one to grasp hand forms. By allowing the video capture and feature extraction processes run in parallel therefore reducing processing delays.

3.2. Gestures: Classification and Recognition

Using data processing after the input feature extraction, a Random Forest classifier exactly marks gestures. Random Forest is an ensemble learning technique with strong and efficient categorization guaranteed by many decision trees.

3.3. Text / Sentence Conversion and Translation

Following classification, the identified gestures become text representations. The device offers multilingual translating so users may translate the identified sign language into English, German, Hindi, Japanese, Tamil, and Telugu. The sentences are formed when user performs same hand sign for 3 second which adds the label in buffer and can be translated.

3.4. Text-to- Speech (TTS) Synthesis

Further processing the translated text using Text to Speech (TTS) technology helps to enable spoken communication. The technology assists those with hearing and speech issues as well as the hearing population to communicate by converting the text into spoken output in the selected language. Operating in a dedicated thread, the voice synthesis mechanism ensures that the system works free from any latency.

3.5. Multithreading for parallel processing

Running various concurrent operations in multithreading helps to improve performance and reduce latency several times. Jobs are divided among threads to optimize efficiency. For gesture detection and video processing, one thread oversees continuous frame capture and feature extraction. Another thread converts gestures into several languages without affecting consciousness. Text to speech synthesis rapidly transforms translated text into speech by means of different thread. Parallel computing allows the system provide a real-time user experience and help to lower latency.

3.6. Effect and Instantaneous Application

For assistive equipment, teaching tools, and regular communication—that is, bridging the deaf-hearing communication gap—the approach is sensible and effective. It is an accessibility tool since it enables low latency, remarkable precision, real-time translation. Multithreading, computer vision, and machine learning help speech and hearing-impaired individuals to interact readily in several languages.

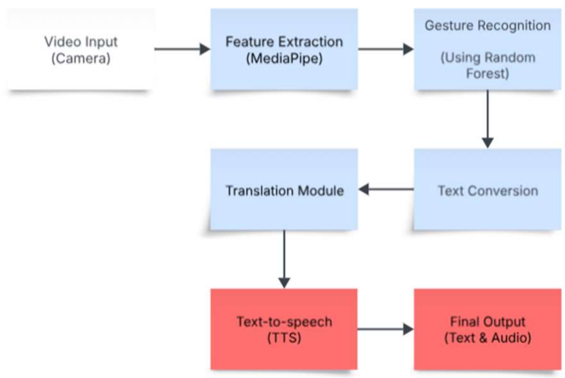

Fig 1.1: Working model

Figure 1.1 shows a real-time sign language recognition and translating system. MediaPipe picks hand and gesture elements from video input coming from a live camera at first. Random Forest a classifier handles these retrieved features for gesture recognition. After that, the acknowledged gesture is turned into text and supported several languages like Hindi, English, Japanese, German, Tamil and Telugu by use of a translating module. A Text-to Speech (TTS) module further handles the translated text to produce both text and audio output, therefore facilitating good multilingual communication.

Figure 1.2 shows the interactive user interface through which user can start camera, stop camera, translate text, play source audio (English), play translated audio, begin / stop sentence generation and clear sentence. It displays status and has dropdown menu to select target language. The live camera feed is at top where user can performs hand sign and will get its dedicated word and translation in a textbox.

3.7 Dataset

We developed a custom dataset using a live camera interface for our Real-Time Sign Language Detection and Recognition Using Live Camera Interface with Multilingual Translation and Speech Synthesis. Our system guarantees balanced representation by capturing three hundred images per gesture. Every gesture is recorded using 150 images of the left hand and 150 images of the right hand, so allowing the model to span many orientations and hand dominance. This procedure is repeated guarantees varied and high-quality data for every word in our dataset. Real-time image acquisition helps us to improve recognition accuracy and flexibility to fit various lighting conditions and hand positions. Starting from this dataset, we train our Random Forest classifier for exact gesture recognition.

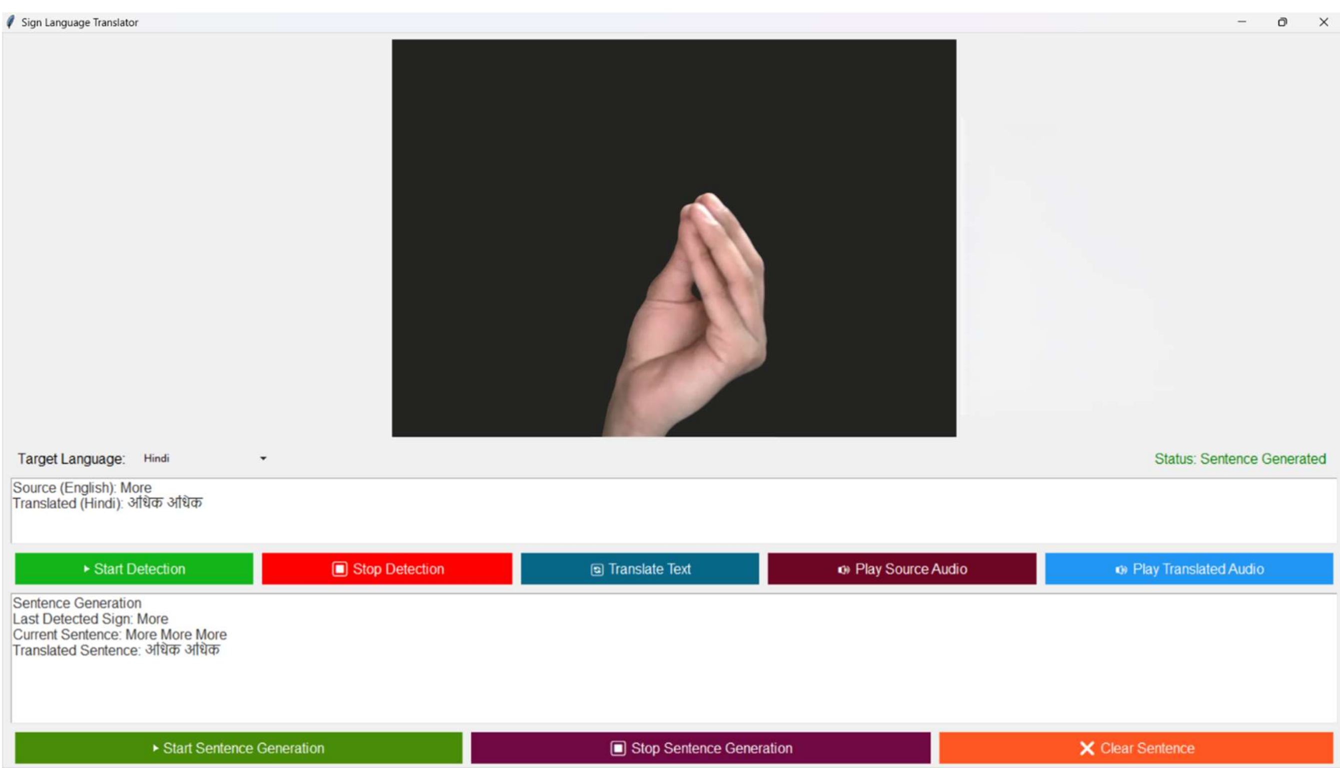

Fig 1.2: User Interface with Sentence generation

IV. MODEL

Random Forest: Designed as an ensemble learning technique, as represented if fig 1.3 the working of Random Forest, which creates numerous decisions trees and compiles their outputs for consistent classification. Ideal for real-time uses, it reduces overfitting and efficiently controls high-dimensional data.

Our project uses Reason Random Forest to help to reduce overfitting and guarantees consistent gesture classification.

- Manages complicated features: hand positions, facial expressions, posture.
- Real-time efficiency offers minimal computational cost rapid inference.
- Strong to Noise: Does great across different backgrounds and lighting.
- Supporting multithreading runs gesture recognition concurrent with speech synthesis and translation.
- Scalable and flexible; can quickly extend to identify novel gestures.
- Interpretable Decisions: Provide obvious routes of decision-making.

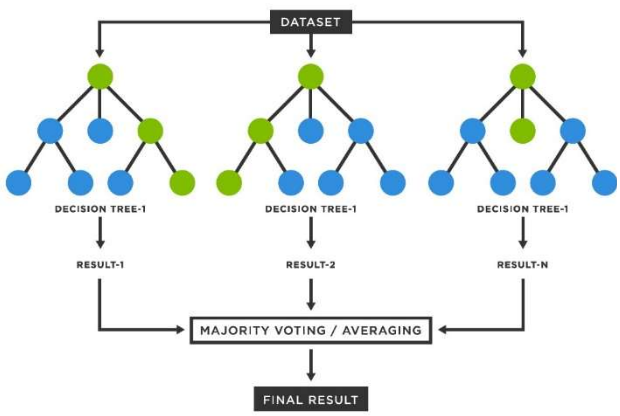

Fig 1.3: Working of Random Forest

Fig 1.3 shows the Random Forest model processes input data through multiple decision trees, each making independent predictions. Their outputs are combined using majority voting or averaging, ensuring higher accuracy, reduced overfitting, and robust final predictions.

Algorithm: Sign language Recognition and translational system

1. Capture real time video from the camera.

2. Using MediaPipe extract hand, face, and postural signals.

3. Organize data preprocessing: Standardize and tidy extracted features.

4. Initialize Random Forest Classifier with N decision trees.

5. Every decision tree will learn on a random sample of the data by means of random feature selection also known as feature bagging.

6. Data point bootstrapping via bootstrapping.

7. Perform Gesture Classification:

- Send the trained Random Forest model real-time extracted features.
- Every decision tree guesses a gesture label.
- Use majority voting to decide on the official recognized gesture.

8. Turn known gestures into text.

9. Translate text into English, German, Hindi, Japanese, Tamil, Telugu.

10. Text to spoken (TTS) generates spoken output.

11. Put multithreading into use to maximize real-time performance:

- First Thread: Gestures for recognition
- Second Thread: Text Translation
- Third Thread: Speech synthesis.

12. Show translated text and run audio output for flawless conversation.

V. RESULT & DISCUSSION

Performance metrics are investigated to assess the efficacy of proposed approach. For this, the following con fusion matrix is used.

Where:

- True Positive (TP): predicted label indicates positive, then accurate the true label.
- True Negative (TN): predicted label indicates negative, then accurate the true label.
- False Positive (FP): predicted label specifies positive, then inaccurate the true label.
- False Negative (FN): predicted label denotes negative, then inaccurate the true label.

F1 Score: Specifically in cases of an imbalance in the distribution of classes, an F1 score is a statistic used to assess the correctness of a classification algorithm. It is the harmonic mean of memory and accuracy, therefore balancing the two measures.

Equations:

Precision: It assesses how many positive labels that anticipated. It measures the accuracy of positive predictions made by a model or system. It's computed following Equation (1):

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

Recall: True positive predictions to the total actual positive's ratio. It's computed following Equation (2):

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

F1 Score: Precision and Recall's harmonic mean is used to compute the F1-score. It's computed following Equation (3):

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

Accuracy: Accuracy refers to the ratio of correctly predicted instances (true positives, true negatives) to the total number of instances in a dataset. Accuracy is analyzed in Equation (4)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Sign	Precision	Recall	F1 score	Accuracy
Hello	0.83	1.00	0.91	0.94
Help	1.00	1.00	1.00	
More	1.00	1.00	1.00	
Sorry	0.76	1.00	0.87	
Yes	1.00	0.93	0.97	

Table 1: Score

Table 1: The table 1 contrasts the performance of models for different words utilizing several evaluation metrics. In F1, Recall, and Precision metrics that assess various dimensions of classification accuracy, such as the identification of true positives the results for all the different words.

VI. CONCLUSION AND FUTURE WORKS

The proposed Real-Time Sign Language Detection and Recognition Using Live Camera Interface with Multilingual Translation and Speech Synthesis effectively bridges the communication gap between hearing and deaf populations. By utilizing computer vision, machine learning (Random Forest), and multithreading, the system ensures very accurate and low-latency gesture recognition. MediaPipe enhances feature extractions; Random Forest provides robust categorization; the system offers multilingual translating into English, German, Hindi, Japanese, Tamil, and Telugu via Text to Speech (TTS), allowing for a large user base from spoken output. This approach supports accessibility, inclusion, and clear communication in practical settings.

Future advancements will specifically incorporate deep learning models including CNNs and LSTMs to increase gesture recognition accuracy even more. The system will expand to include new regional and international languages and adapt to mark continuous sign language sentences instead of single gestures.

Using edge devices like Raspberry Pi and AI-powered wearables for real-time usability will help to investigate hardware efficiency. Furthermore, given top priority will be dataset expansion to increase model generalization in several contexts. The creation of a specific mobile and web application to guarantee the system may be extensively adopted in daily communication and assistive technologies and so increase its accessibility and user-friendliness. These developments will help define the system as a scalable, effective, inclusive tool for people with speech and hearing disabilities.

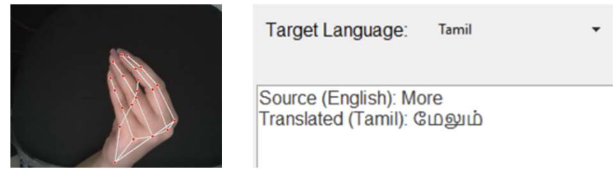

Fig 1.4: Example 1

Fig 1.5: Example 2

Fig 1.4 & 1.5 shows how our model works. In fig 1.4 the sign shows More which is converted to English and Tamil language which the user selected. In fig 1.5 the sign is Hello which is converted to Tamil.

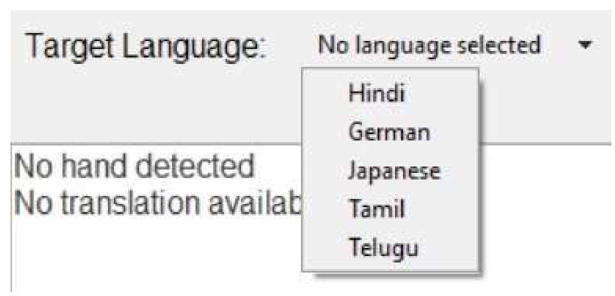

Fig 1.6: Language menu

Fig 1.6 shows the dropdown menu through which the user can select the desired language like Hindi, German, Japanese, Tamil and Telugu.

REFERENCES

[1] Núñez-Marcos, Adrián, "A survey on Sign Language Machine Translation," Expert Systems with Applications, Vol. 213, 2023.Sakshi Bhangale, Kunal Deokar , Translation: Code-Mixed Language (Hinglish) to English, IJARSCT,2022

[2] Lianyu Hu, "Scalable Frame Resolution for Efficient Continuous Sign Language Recognition," Pattern Recognition, Vol. 145, 2024.Nicole D'Souza and Devarsh Patel, Survey on Hinglish to English Translation and Classification Techniques , Recent and Innovation Trends in Computing and Communication, 2023

[3] Bragg Danielle, "Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective," ASSETS, 2019.Levi Corallo and Guanghui Li , A Framework for German- English Machine Translation with GRU RNN , Published in the

Workshop Proceedings of the EDBT/ICDT 2022 Joint Conference (March 29-April 1, 2022), Edinburgh, UK

[4] Suharjito, Ricky Anderson, "Sign Language Recognition Application Systems for Deaf-Mute People," ICCSCI, 2017.

[5] Rao, G. Mallikarjuna, "Sign Language Recognition using LSTM and MediaPipe," ICICCS, 2023.

[6] Gan, Shiwei, Jiang, Zhiwei, "Contrastive Learning for Sign Language Recognition and Translation," IJCAI, 2023.

[7] Angelova, Galina, "Using Neural Machine Translation Methods for Sign Language Translation," Student Research Workshop, 2022.

[8] Camgoz, Necati Cihan, "Neural Sign Language Translation," IEEE Conference, 2018.

[9] Chen, Yutong, "A Simple Multi-Modality Transfer Learning Baseline for Sign Language Translation," IEEE/CVF Conference, 2022.

[10] Camgoz, Necati Cihan, "Neural Sign Language Translation," IEEE Conference, 2018.

[11] Chen, Yutong, "A Simple Multi-Modality Transfer Learning Baseline for Sign Language Translation," IEEE/CVF Conference, 2022.

[12] Rastgoo, Razieh, "Sign Language Production: A Review," IEEE/CVF Conference, 2021.

[13] Ajay, S., "Indian Sign Language Recognition Using Random Forest Classifier," IEEE (CONECCT), 2021.

[14] Sunanda das , "A Hybrid Approach for Bangla Sign Language Recognition Using Deep Transfer Learning Model with Random Forest Classifier," Elsevier, 2022.

[15] Suharjito, Ricky Anderson , “sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output” , 2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI.