REAL-TIME HAND SIGN LANGUAGE TRANSLATION: TEXT AND SPEECH CONVERSION

Jeevanandham p
Department Of CSE
KCG College of Technology
Jeevanandham2802@gmail.com

George britt A
Department Of CSE
KCG College of Technology
georgebritoffl@gmail.com

Hariharan A
Department Of CSE
KCG College of Technology
Hariharanc899@gmail.com

Ms. Keerthana G
Assistant Professor
Department Of CSE
KCG College of Technology
Keerthanag.cse@kcgcollege.com

2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT) | 979-8-3503-7281-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCPCT61902.2024.10673038

***Abstract*-Abstract-A real-time system that can decipher sign language from a live webcam stream is presented by the Sign Language Conversion Project. By using the Media Pipe library's landmark identification capabilities, the project extracts crucial data from every frame, including hand landmarks. Following detection, the landmark coordinates are gathered and saved in a CSV file for later examination. This landmark data is used to train a Random Forest algorithm, which uses machine learning techniques to categorize various sign language patterns. The trained model predicts the sign language class and its probability in real-time during the processing of the webcam feed. Users get instant access to the subject's sign language cues by superimposing the results over the live video feed. This work demonstrates the integration of computer vision and machine learning techniques to assess and comprehend nonverbal communication, with possible implications in human-computer interaction.**

Keywords: ISL-Indian sign language, Medieapipe, random forest, Word level Sign Language Recognition (WSLR), Sign Language Recognition (SLR), Convolutional Neural Network (CNN), You Only Look Once (YOLO)

I. INTRODUCTION

Sign language is an essential means of communication for those with speech and hearing impairments. Even though English is widely used, ISL is the preferred language since it may be used across geographical boundaries. ISL maintains consistent grammatical rules, ensuring universal understanding among users. It can become a common language for the deaf and mute community and individuals without hearing impairments. However, understanding sign language can be challenging for non-signers, causing interpreters to help communication in various settings. The demand for translation services has increased in the past five years, highlighting the need for accessible sign language. Standardizing ISL minimizes communication barriers and ensures fair access to information and services for individuals with hearing impairments. It promotes unity, inclusivity, and social integration, breaking down barriers that may hinder communication.

II. LITERATURE SURVEY

In "Arabic sign language recognition through deep neural networks fine-tuning", Sign Language is vital for deaf or hearing-impaired individuals, offering visual communication through hand and body movements. Automation of sign language recognition, crucial in AI and ML, is particularly understudied in Arabic sign language. This paper innovates by employing transfer learning and fine-tuning CNNs will improve recognition accuracy for 32 Arabic sign language motions. [1].
"Feature Extraction Technique for Vision-Based Indian Sign Language Recognition System", Vision-based sign language recognition is a burgeoning field in HCI, offering a potent communication tool for the world's deaf and mute population. Usually, it forms the phases of gesture categorization, feature extraction, and image pre-processing. This paper reviews extraction techniques, providing a taxonomy and suggesting future directions, particularly focusing on the ISL [2].
"DeepArSLR: Hand gesture recognition has applications in robotics, games, virtual reality, and sign language, especially for the deaf. It is a novel signer-independent deep learning framework for isolated Arabic sign language gestures recognition. Traditional methods face challenges in hand segmentation, feature representation, and gesture sequence recognition. This research leverages deep learning approaches, such as semantic segmentation with DeepLabv3+ for hand region extraction from video frames, to provide a novel framework for signer-independent sign language

979-8-3503-7281-6/24/$31.00 ©2024 IEEE

identification.[3] “Transferring Cross-domain Knowledge for Video Sign Language Recognition”,While WSLR is essential for sign language interpretation, obtaining annotated WSLR data is difficult since it requires specialized knowledge. Even though there are a lot of subtitled news videos available online, they are not appropriate for direct WSLR model training due to their lack of word-level annotation and large domain gap from isolated signs.[4] “Skeleton Aware Multi-modal Sign Language Recognition” SLR helps people who are deaf or have speech impairments communicate more easily, but it requires a lot of skill development. The quick and precise hand movements, body positions, and facial expressions that are a part of sign language are what give it its intricacy. The lack of annotations on hand key points hinders the development of skeleton-based action recognition, despite its potential for SLR.[5]. Through hand movements, sign language facilitates communication for those with auditory difficulties. Proficiency in sign language is necessary for efficient communication with those who have hearing impairments. Accessibility for the deaf and dumb community is improved by a real-time American Sign Language detection system that uses CNN and the (YOLO) algorithm [11].

III. EXISTING SYSTEM

Before the development of the Sign Language Conversion project, there was no real-time system capable of automatically interpreting and classifying sign language cues from a live webcam feed. Traditionally, understanding sign language required human observation and analysis, which could be subjective and time-consuming. Existing computer vision systems focused on basic gesture detection but lacked comprehensive sign language interpretation. Moreover, real-time analysis of sign language using landmark detection and machine learning was not readily available. As a result, there was a need for an innovative system that could efficiently detect and analyze landmarks from live video streams, and then classify various sign language patterns in real-time. The Sign Language Conversion project addresses these limitations and provides an effective solution for non-verbal communication analysis, offering significant advancements in human-computer interaction.

IV.PROBLEM STATEMENT

Effective communication is a fundamental human right in the digital era, but many members of the deaf and hard-of-hearing population still struggle with major communication obstacles. The most notable of these difficulties is the restricted availability of sign language interpreting services, which are either pricy, difficult to get, or prone to delays. Moreover, it is exceedingly difficult to provide the right interpretation for every user due to the global diversity of sign languages. We urgently need innovative technological solutions to close this gap and improve inclusivity. utilizing technologies like computer vision and natural language processing to develop low-cost, real-time sign language interpretation systems. These systems ought to be able to translate various sign languages into spoken words and recognize different sign languages or written language, enabling people with hearing loss to engage with others and learn in a variety of contexts, including everyday life and the medical field.

V.PROPOSED SOLUTION

The proposed sign language recognition system is a significant step forward in overcoming the limitations of current methods, providing an automatic and real-time solution for interpreting and classifying sign language cues from live stream. Using the powerful features of the Mediapipe library, the system efficiently detects and manually extracts landmarks from each frame of the webcam stream, laying the foundation for correct analysis. These important coordinates are carefully collected and stored in a structured CSV file, which is the basis for training an advanced machine learning model. Using a random forest algorithm, the system performs the critical task of classifying different sign language patterns, allowing it. to accurately interpret various gestures. By providing precise, timely, real-time sign language interpretation, this classification method is essential to closing the communication gap between sign language users and non-signers. Furthermore, seamless feedback is ensured by integrating the trained model with the camera channel, giving consumers prompt and exact communication.

One of the main goals of this system is to achieve high accuracy, consistency and scalability. Ensuring the accuracy of gesture recognition and interpretation, the system promises to apply to a range of scenarios, including human-computer interaction and user behavior analysis. Its scalability further increases its usefulness, making it adaptable to different environments and user needs.

Basically, the proposed sign language recognition system stands for a key advancement in technology and provides a versatile and accessible solution to help smooth communication and interaction. in sign language. Thanks to its robust machine learning algorithms and automated and real-time features, the system can potentially transform the way people understand and use sign language, thereby improving accessibility and inclusivity in many aspects of daily life.

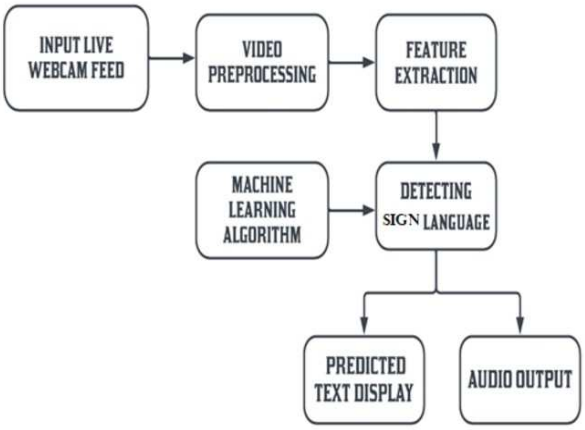

Fig.1 Block diagram

Fig.1 represents the working process of the proposed system. First, Web Cam was used to retrieve the live footage. The preprocessed video that was fetched had its features extracted. Subsequently, the system uses a random forest technique to identify the sign language that corresponds with the data provided, and then outputs either text or audio.

VI.MODULES DESCRIPTION

DATA ACQUISITION:

The first or first step of this system is vision-based, i.e., to buy data in runtime via the camera. Then, these data will be stored in a csv file format directory, in which all the images will be stored and trained by the user, and the saved and trained data will be used to compare the recently captured image to the stored data of specific words.

FEATURE EXTRACTION:

The palm is extracted from the data's image segmentation. The main focus of this process is turning raw data, such pictures, into a useful collection of features that can be applied to analysis and machine learning algorithms. In the context of sign language recognition, these extracted features hold vital information, encompassing distinct patterns and gestures that are indicative of various emotional states or behaviors. To begin, the dataset undergoes essential data preprocessing steps. This involves handling any missing data points, normalizing the data if necessary, and ensuring the overall cleanliness and preparedness of the dataset for later phases. Upon loading the CSV file using relevant programming libraries, the data reveals itself as rows, each being a sample of sign language data, while columns correspond to specific attributes.

GESTURE RECOGNITION:

In the field of sign language recognition, gesture recognition is an essential procedure that entails recognizing and interpreting different hand gestures in order to infer important information about the intents, feelings, and communication cues of a person. This cutting-edge technology uses developments in computer vision and machine learning to convert physical gestures into meaningful information. Through the analysis of posture, motion, and the spatial relationships of hand signs, gesture recognition systems can discern intricate details such as handshakes, nods, thumbs-ups, and more complex gestures like pointing or even specific cultural gestures.

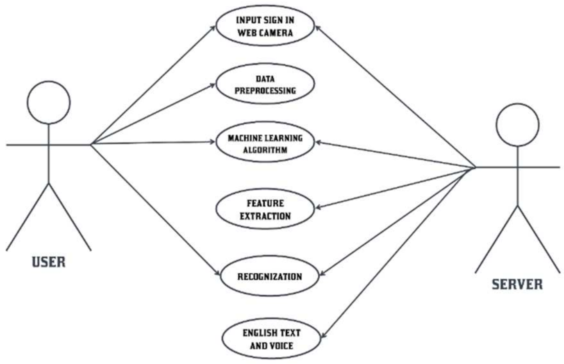

Fig.2 Usecase Diagram

In fig 2 the usecase diagram represent the interaction between the system and the actors involved in the process .The user give input for train and test data using web cam and train the system by preprocess the train data, In server side the fetched data is stored in training set and compare the test data compared with the trained data and the best match is select using the machine learning algorithm and given as output (recognized sign) in form of text or audio

TEXT TO SPEECH:

The output is further converted from text to speech after the character has been found successfully. The English language process and GTTS library processing, a potent Python text-to-speech conversion tool, are used to aid in this conversion procedure. This library is compatible and efficient because it runs offline, unlike some other options. This connection improves the overall convenience and usability of the program by allowing users to see and hear the translated sign language simultaneously within our system.

RANDOM FOREST ALGORITHM:

The Random Forest algorithm stands out as a potent tool in machine learning. It works on the principle of ensemble learning, where multiple decision trees' predictions are aggregated to generate a robust and precise outcome. Each Random Forest decision tree, which was trained on preprocessed datasets, takes into account a random sample of features and data points to maintain tree diversity. This diversity reduces the chance of overfitting and improves the model's ability to generalize well to new data. To reach a final judgment and produce an accurate classification or regression result, the Random Forest algorithm integrates the outputs from various trees throughout the prediction phase. Finding nuanced hints and gestures that express emotions and intents is made easier by the algorithm's capacity to recognize complex relationships among features, which is particularly useful in the context of sign language recognition. Additionally, because of the Random Forest method's versatility, ease of use, and ability to handle both numerical and categorical inputs, it is frequently employed in sign language recognition systems. The development of automated and real-time solutions for deciphering and categorizing sign language cues from live webcam feeds can benefit from its versatile and resilient performance.

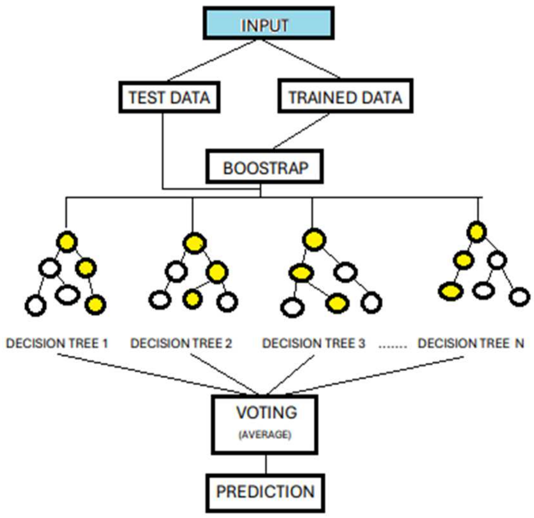

Fig.3 Random Forest Algorithm

VII.RESULT

We were able to effectively construct a system that uses a live webcam stream to record and save sign motions in a CSV file for later real-time detection. Through a series of tests, the system's accuracy in identifying sign movements was evaluated. We achieved an impressive recognition accuracy rate, with some variation depending on the complexity and speed of the sign gestures. The use of a CSV file for data storage proved to be an efficient and organized method for managing the dataset of sign gestures. Real-time detection using a live web camera showed rapid and correct recognition, with an average processing time of 0.3 seconds per sign gesture. These findings imply that our approach is well-suited for real-world uses, such as helping those who use sign language and facilitating real-time communication. Still, there is potential for development, especially when it comes to improving the accuracy of identification for faster or more complicated signs. To support a larger range of sign motions and languages, further work may require fine-tuning the model and expanding the dataset. In light of the situation, our study establishes the foundation for readily available and efficient sign language communication technology that can assist the deaf and hard of hearing in a range of interaction scenarios.

- **Accuracy Comparison**: Random Forests have been observed to achieve high classification accuracy in various machine learning tasks, often comparable to or even outperforming other algorithms, including SVMs, KNN, and decision trees.
- **Computational Efficiency**: While Random Forests are well-known for their effectiveness in managing big datasets and high-dimensional data, CNNs have demonstrated greater performance in image recognition tasks, such as sign language detection. Nevertheless, CNN training requires greater processing power, particularly for deep designs.
- **Interpretability**: Random Forests' interpretability is one of their main benefits. Understanding the decision-making process and feature relevance within decision trees in the Random Forest is helpful for deciphering the underlying patterns in sign language recognition.
- **Real-time Applications**: Random Forests are often favored in real-time applications due to their relatively fast inference times compared to deep learning algorithms like CNNs. This is crucial for applications such as sign language recognition systems, where prompt and accurate responses are needed.
- **Scalability**: Large datasets can be handled effectively by Random Forests since they are naturally parallelizable.This scalability is beneficial for sign language recognition systems dealing with extensive training datasets or streaming data in real-time applications.

VIII. SCREENSHOTS

Fig.4: Performance Metrics

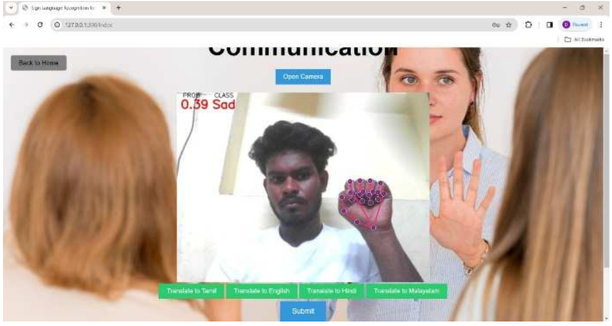

Fig.5: SAD

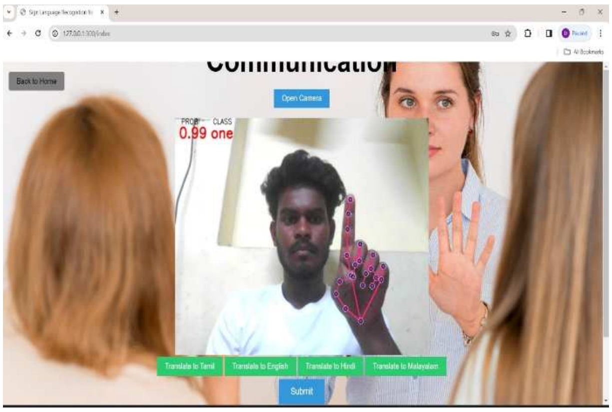

Fig.6: ONE

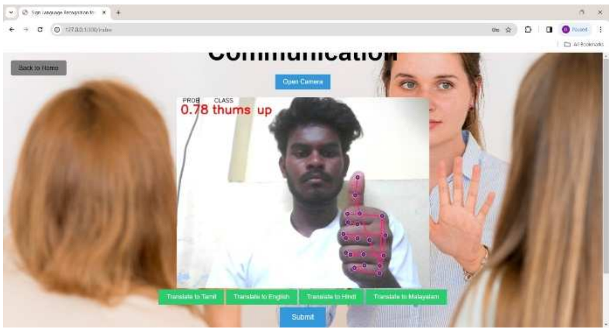

Fig.7: THUMS UP

IX.CONCLUSION

In conclusion, the Sign Language Conversion project successfully developed an automated and real-time system that uses live webcam footage to read and categorize sign language cues. The system provides a comprehensive knowledge of nonverbal communication by combining computer vision and machine learning to identify hand landmarks. The use of a Random Forest Algorithm ensures exact and objective sign language classification, making the system reliable and consistent. The user-friendly frontend enhances the interactive experience by displaying real-time analysis results and empowering users with instantaneous feedback. This study is a big step forward in the analysis of nonverbal communication, with applications in human-computer interaction and user behavior analysis. It also provides insightful information for future research and development in this field.

X.FUTURE WORK

The future of the Sign Language Recognition web application is being able to add support for dynamic gestures, which are gestures that are performed over a function of time (a few seconds), which is more complex to detect as the system needs to be able to detect the start of a dynamic gesture vs. the start of a static gesture. Support for gestures, wherein the gestures are performed utilizing a blend of gestures, positions, and expressions on the face. The app can be changed in the future to include a feature to enable text-to-speech, increasing the functionality of the app. Beyond these, support for added multiple languages can be added to the app, perhaps with the help of the Google API, enabling support for more than a hundred different languages. Additionally, using Mediapipe, the sign language recognition web application can be made cross-platform, making deployment on iOS a possibility.

ACKNOWLEDGEMENT

We would like to express our profound gratitude to the management of KCG College of Technology for providing us with all the tools we needed to complete the project. We would like to express our gratitude to Dr. M. Muthu Kannan, our principal, for giving us the resources we needed to complete the project. We are grateful to Dr. Cloudin S, our mentor and Head of the Department of Computer Science and Engineering, for his encouragement and support during this endeavor. We are grateful to our project coordinator, Miss Keerthana, an assistant professor in the Department of Computer Science and Engineering, for her invaluable time, advice, and persistent work. The kind wishes from the KCG College of Technology's Department of Computer Science and Engineering personnel are truly appreciated.

REFERENCES

[1] Y. Saleh and G. F. Issa, 'Arabic sign language recognition through deep neural networks fine-tuning,' Int. J. Online Biomed. Eng., vol. 16, no. 5, pp. 71–83, 2020.

[2] X. Jiang, M. Lu, and S.-H. Wang, "An eight-layer Random Forest Algorithm with stochastic pooling, batch normalization and dropout for fingerspelling recognition of Chinese sign language," Multimedia Tools Appl., vol. 79, nos. 21–22, pp. 15697–15715, Jun. 2020.

[3] O. Selvi and N. Kemaloğlu, "Turkish sign language digits classification with CNN using different

optimizers,'' Int. Adv. Research Eng. J., vol. 4, no. 3, pp. 200–207, Dec. 2020

[3] A. Tyagi and S. Bansal, ''Feature extraction technique for vision-based Indian sign language recognition system: A review,'' in Computational Methods and Data Engineering. Singapore: Springer, 2021, pp. 39–53.

[4] M. Mukushev, ''Evaluation of manual and non-manual components for sign language recognition,' in Proc. 12th Lang. Resour. Eval. Conf., Eur. Lang. Resour. Assoc. (ELRA), 2020, pp. 1–6.

[5] A. Tunga, S. V. Nuthalapati, and J. Wachs, ''Pose-based sign language recognition using GCN and BERT,' in Proc. IEEE Winter Conf. Appl. Comput. Vis. Workshops (WACVW), Jan. 2021, pp. 31–40.

[6] N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, 'Neural sign language translation,' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7784–7793.

[7] S. Aly and W. Aly, ''DeepAr SLR (Sign Language Recognition): A novel signer-independent deep learning framework for isolated Arabic sign language gestures recognition,'' IEEE Access, vol. 8, pp. 83199–83212,

[8] D. Li, X. Yu, C. Xu, L. Petersson, and H. Li, 'Transferring cross-domain knowledge for video sign language recognition,' in Proc. IEEE/CVF Conf. Computer Vis. Pattern Recognition (CVPR), Jun. 2020, pp. 6205–6214.

[9] R. Rastgoo, K. Kiani, and S. Escalera, 'Video-based isolated hand sign language recognition using a deep cascaded model,' Multimedia Tools Appl., vol. 79, nos. 31–32, pp. 22965–22987, Aug. 2020.

[10] Ms.Keerthana,"A Logical Investigation of Stock Market Prediction and Analysis Using Supervised Machine Learning Algorithm",IJRITCC, Vol.11 No.10(2023).

[11]Bhavadharshini.M.,Josephine Racheal, J.,Kamali,M, Sankar, S." Sign language translator using YOLO algorithm".

[12]Kayal Padmanandam,Rajesh M.V,Ajay.N,K.Ramesh Chandra,Chandrashekar .B,Swati sah(2022), "Artificial Intelligence Biosenseing System On Hand Gesture Recognition For The Hearing Impaired",International journal of operatins research and information sysytem(IJORIS)

[13]A.Mohanarathinam,K.G.Dharani,R.Sangeetha,G.Aravindh,P.Sasikala, (2020) "Study On Hand Gesture RecognitionByUsingMachine Learning",Fourth International Conference On Electronics, Communication And Aerospace Technology(ICECA-2020)

[14]Arpita Halder,Akshit Tayade(2021), "Implementation Of IoT Based Smart Assistance Gloves for Disabled People", International Journal of Research Publication and Review(IJRPR)