IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

An Intelligent Real-Time Sign Language Recognition System Using YOLOv5 and MediaPipe for Enhanced Human–Computer Interaction

[1]Meghana B K, [2]Dhanyashree A L

[1]Bachelor of Computer Applications, [2]Bachelor of Computer Applications

[1]ASC Degree College, [2]ASC Degree College Bengaluru, Karnataka, India

Abstract:

For years in Western culture, sign language has been an ancient visual language and method of communicating. It consists of the use of traditional signs, gestures or a manual alphabet using your hands to assist with signing. The Machine Learning-based sign language identification system is a project that allows someone who has a disability to communicate and provide an output from sign language into text or audio; therefore, the machine can read the signs, and the machine will know what the user is trying to say even if the user has no knowledge/sign language. To provide better services to disabled individuals is the purpose of this project through advanced technology. The purpose of Sign Language is to communicate the user's need for information. The Project will capture sign language using specific technologies and will provide an audio notification of its result on the website. The Project is developed using Machine Learning technology and other advanced technologies. The purpose of the Project is to use hand sensors used to detect sign/capture sign as input to identify the needs of the disabled person. The purpose of the project is to provide a path for helping the world through humans.

Keywords: Gesture navigation, TensorFlow, Keras, OpenCV-Python, matplotlib, pandas, numpy, MediaPipe, seaborn, Flask, scikit-learn, pyttsx3, YOLOv5.

I. INTRODUCTION

For those who have hearing or speaking disabilities, sign language is one of the most frequently used ways of communicating visually. Millions of individuals around the world use sign language as a main method of communicating with others. Yet many users of sign language encounter problems with being able to communicate effectively due to a lack of knowledge by individuals who do not understand sign language. The continued existence of communication obstacles, particularly in terms of education, healthcare, employment, and public services, highlights the need for intelligent assistive technology that will automatically convert sign language into verbal words.

Recent advancements in Artificial Intelligence (AI), Computer Vision and Deep Learning technologies have provided systems that can now recognize sign language gestures with much greater accuracy than was previously possible. A typical SLR system involves capturing hand gestures with a camera, pre-processing the input images, extracting the relevant properties of the hand, and then classifying the hand gesture into a corresponding word or speech. Although significant progress has been made with current SLR systems, many of these systems still experience problems recognizing hand signs due to factors such as lighting conditions; background complexity; whether or not the individual's hand is occluded; and individual variations, including the appearance and physical characteristics of individuals, ultimately leading to decreased accuracy of hand sign recognition and limiting the use of these systems in the real-time application.

To overcome these issues, this study proposes a smart visual-based sign language recognition system that utilizes YOLOv5 for precise hand detection, MediaPipe to detect hand landmarks, and TensorFlow and Keras to classify gestures. OpenCV is used for processing images, and Flask provides a web interface that can handle both uploaded images and live webcam feeds. The recognized gestures are presented as text and are also convertible to speech so that communication can occur efficiently and in real-time. Overall, the proposed system provides a solution that is accurate, efficient, and easy to use; it provides improved access to people with disabilities who are unable to hear or speak, and contributes towards the development of intelligent assistive technology.

II. METHODOLOGY

The proposed sign language recognition (SLR) system processes hand gestures sequentially through the following five steps:

2.1 Hand Gesture Acquisition

The SLR system captures images of hand gestures using either a web camera or an uploaded file. These images of hand gestures will be the input for the recognition process.

2.2 Hand Gesture Image Preprocessing

In order to increase the quality and consistency of the hand gesture images, the images must be processed prior to recognition. The preprocessing will include the removal of noise, resizing of images, normalisation, and the segmentation of the hand from the background image so that it can be extracted for feature extraction.

2.3 Model Training

The recognition model will be trained using a dataset that has been preprocessed. YOLOv5 is used to help identify the hands, while MediaPipe provides coordinates of the hand landmark features. TensorFlow and Keras are then used to learn the patterns of gestures and classify different sign language gestures.

2.4 Model Testing

In order to verify the accuracy of the hand gesture recognition system, the trained model is tested using a set of unseen images of hand gestures. The overall recognition capabilities of the system will be evaluated based on its ability to identify hand gestures in different environments.

2.5 Performance Evaluation

The performance of the hand gesture recognition system will be evaluated by calculating four separate evaluation metrics: accuracy, precision, recall and F1 score. These metrics will provide an indication of the reliability and success of the proposed system.

2.6 Deployment

A Flask web application incorporates the trained recognition model in order to allow users to identify sign language gestures. Users may provide an image to identify a sign gesture or, if they have access to a camera, they may use that method to identify sign gestures through the web application. The web application will display the identified sign gesture (as text); this gesture can also be converted into spoken language to improve communication between all parties involved.

III. REQUIREMENTS

3.1 Functional Requirements

User Interaction: Provide a simple and user-friendly web interface for live interaction. The system allows users to upload either still image(s) of hand gestures or to capture hand gestures using a webcam. Display the interpreted (recognized) sign as text and optionally convert it to speech.

Image Acquisition: Images of gestures should be captured live through a webcam or uploaded images. The immediacy of acquisition should be real-time with little or no delay.

Pre-processed Image: Images must be pre-processed to improve quality by resizing, normalising, and removing noise. Images must have their hands segmented to improve the accuracy of recognition.

Hand Detection and Extraction of Landmarks: To determine the hand regions (ie. to determine where to apply finger movement measurements) to detect hand region detectors, use the YOLOv5 Object Detection Model. For hands that are identified, create 21 landmarks (i.e., coordinate points of hand location), using MediaPipe appropriate for gesture movements.

Gesture Recognition: A gesture will be classified utilising an already trained Internal Neural Network (which will use learners from TensorFlow & Keras). With learned models, the same gesture will be achieved with high accuracy.

Generating Output: Display the identified gesture as text on the web application. Convert the identified text to speech output to enhance accessibility.

Managing Models: Incorporate the trained model with the web application to load during the application startup. There should be a pathway to retrain and/or incorporate additional sign language datasets into the application.

3.2 Non-Functional Requirements

Performance: Recognize hand gestures in real-time with minimal delay and predict quickly from images received via upload and live tracking of video from a camera.

Accuracy: Achieve a high level of accuracy when recognizing hands in different positions, light settings, and backgrounds.

Scalability: Facilitate future development for recognizing additional sign language gestures and use larger datasets.

Reliability: Ensure users receive similar predictions with few errors for extended periods of use.

Security: Safeguard user uploaded photographs as well as user data when processing images. Ensure a secure means for communicating back and forth between client and web application.

Maintainability: Design using a modular architecture to facilitate future updates, debugging operations and the addition of new features.

Usability: Create user interfaces known for their ease of use, responsiveness and intuitive functionality, as well as providing functionality to accommodate varying levels of technical expertise from the users.

Compatibility: Support for the most up-to-date (current) web browsers on the majority of personal computers without any special hardware is required.

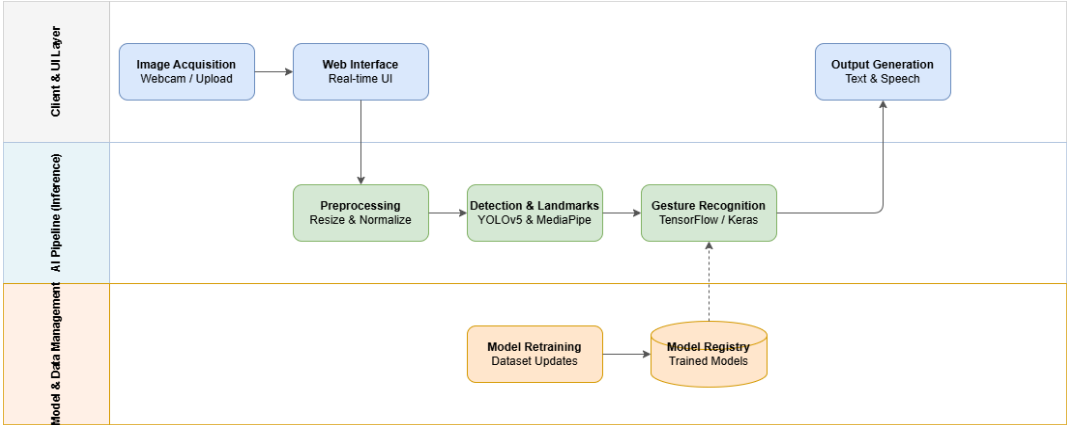

Figure 1: Functional Architecture of the Proposed Sign Language Recognition Framework.

IV. SYSTEM ARCHITECTURE

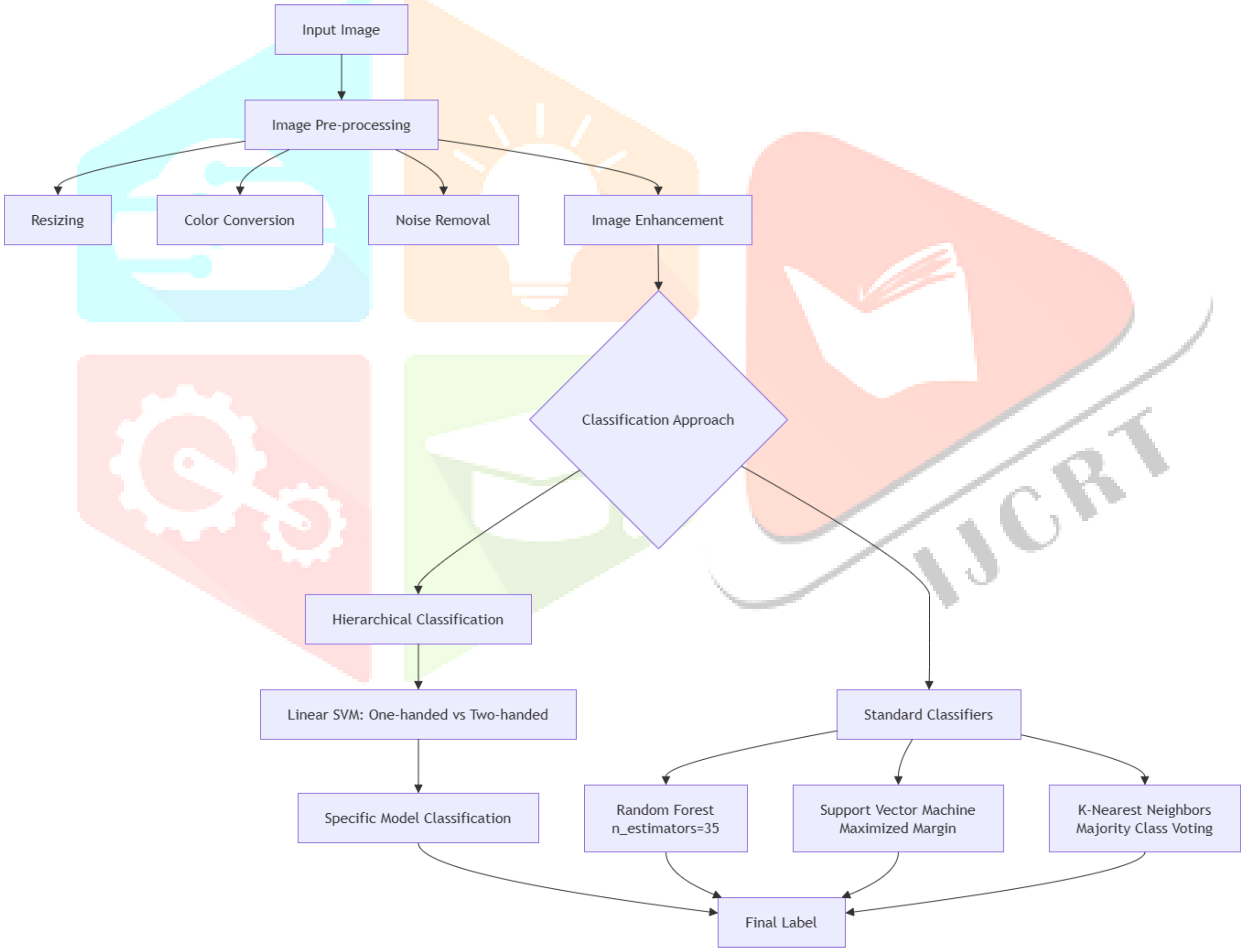

Figure 2: Functional Architecture of the Proposed Sign Language Recognition Framework.

Image Pre-Processing Techniques: Applying several types of pre-processing techniques on an input image allows for successfully reducing unwanted noise and enhancing image quality. Techniques that can be applied to an input image include resizing (scaling), converting colours, and reducing unwanted noise through various combinations of all three. The accuracy of your output will largely depend on how well you select the correct types of pre-processing techniques.

Image Enhancement Techniques: Improving an image's visual quality is a fundamental aspect of image processing. Improving an image visually and therefore increasing the quality of an image will positively impact human or machine analysis of that image. Improving and restoring an image's visual quality is vital to the success of the image processing process.

Hierarchical Classification: One of the main reasons that the accuracy has been lower than expected is because there are so many classes available. So we attempted to break up the classification problem into a number of levels, or as hierarchical classification, instead. First, a linear kernel SVM model will classify the alphabets as either one-handed or two-handed. In addition, the alphabet first needs to be classified into either one-handed or two-handed before it can then be sent into the correct model and assigned an appropriate label depending upon how it has been classified.

Random forest: Is a well-known algorithm for its ability to deal with noisy data and provide estimates of feature importance. As a result, it has become a very popular machine learning algorithm for many application areas in the real world. A random forest is an ensemble (committee) of multiple decision trees that generates predictions for problems involving regression or classification, and then aggregates the predictions generated by all of the decision trees in the forest.

Support vector machine: A Support Vector Machine (SVM) is a machine learning algorithm that can be used in sign language recognition. This type of machine learning algorithm uses a mathematical model to identify the boundary separating two classes in a dataset and will find the boundary with the largest margin (distance from the boundary to the closest point in each class). The data points that are closest to the boundary will be the ones that have the greatest influence on the creation of that boundary (support vectors). SVMs are flexible classifiers that can create linear (line) or non-linear (curve) decision boundaries when distinguishing between two classes and can also work with datasets that have a high number of dimensions (the number of features) in the data. There are many features of sign language to consider, including hand shape, the way the hands are held, and how the hands move.

K-NEAREST NEIGHBOUR: A K-Nearest Neighbour (KNN) Algorithm is an algorithm that is also used for recognising sign language through machine learning. KNN's methodology works by classifying a point in the dataset of a variable type (dependent variable or independent variable) to its nearest neighbours and finding the most frequent class in this specified number (k). The k is something that you can adjust and change, depending on how well it performs. Once you have determined the k value that gives you the best performance when you have created your test data using KNN for gesture recognition, it can be classified by matching it against your test data (using sign language gestures) by comparing hand shape, movement, and direction. The KNN algorithm is relatively simple and easy to use, but it may require quite a bit of time to compute on a large set of data, and it has limitations for working with higher-dimensional datasets.

IV. WORKFLOW

The sign language recognition system outlined in this document begins with the acquisition of hand gesture images either via a live webcam feed or uploaded from other sources. The acquired images will be processed first for resizing and normalizing and then for noise removal through various methods in order to increase their quality. Next, the processed image will be passed from the preprocessing stage to the YOLOv5 model in order to detect the hand, which will then pass through MediaPipe to extract the relevant points representing hand landmarks. The features extracted from the image will then be classified using a deep learning model developed using TensorFlow and Keras in order to identify the correct sign language gesture. The final step of the recognition process is to convert the recognised gesture to text and then convert it to audio speech through the web application, allowing for real-time communication by persons with hearing and/or speech disabilities.

Figure 3: Home page of the proposed Sign Language Recognition web application.

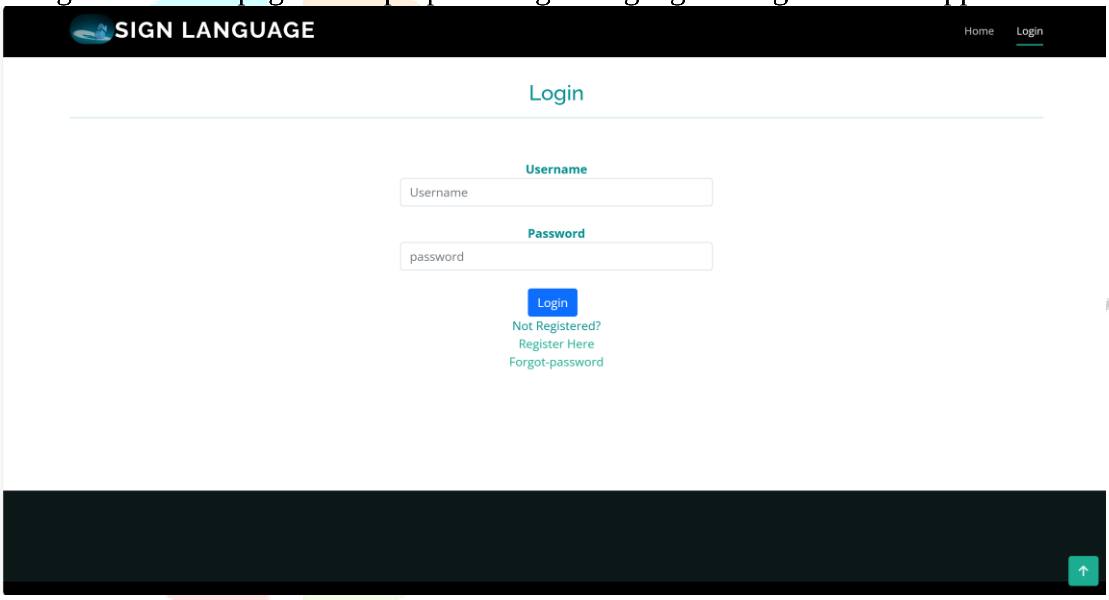

Figure 4: User login interface for authenticated access to the Sign Language Recognition system.

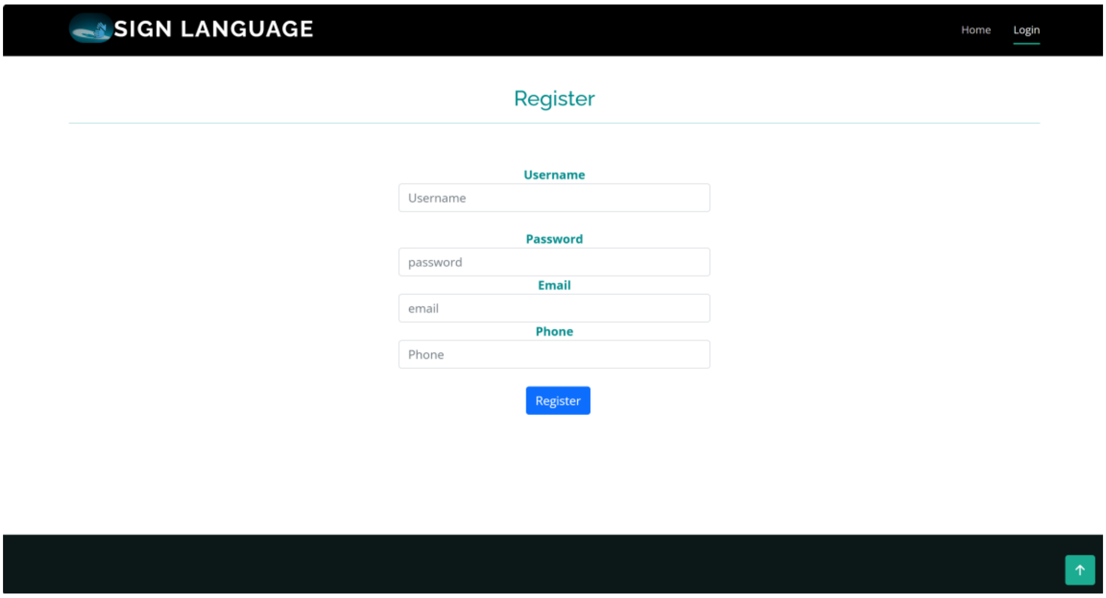

Figure 5: User registration interface for creating a new account in the Sign Language Recognition system.

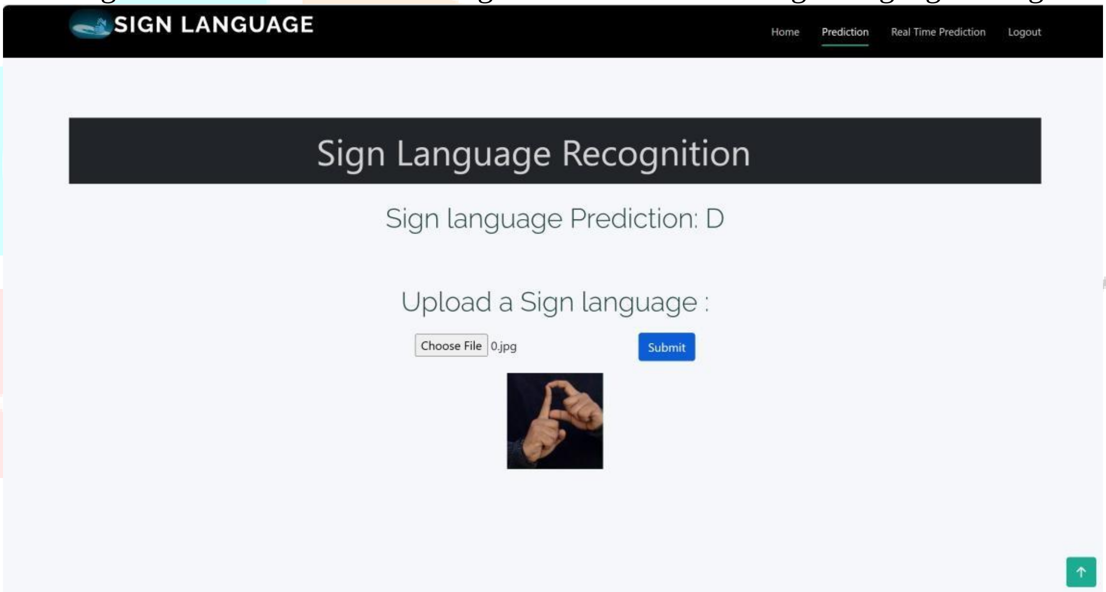

Figure 6: Image-based sign language prediction using the proposed recognition model.

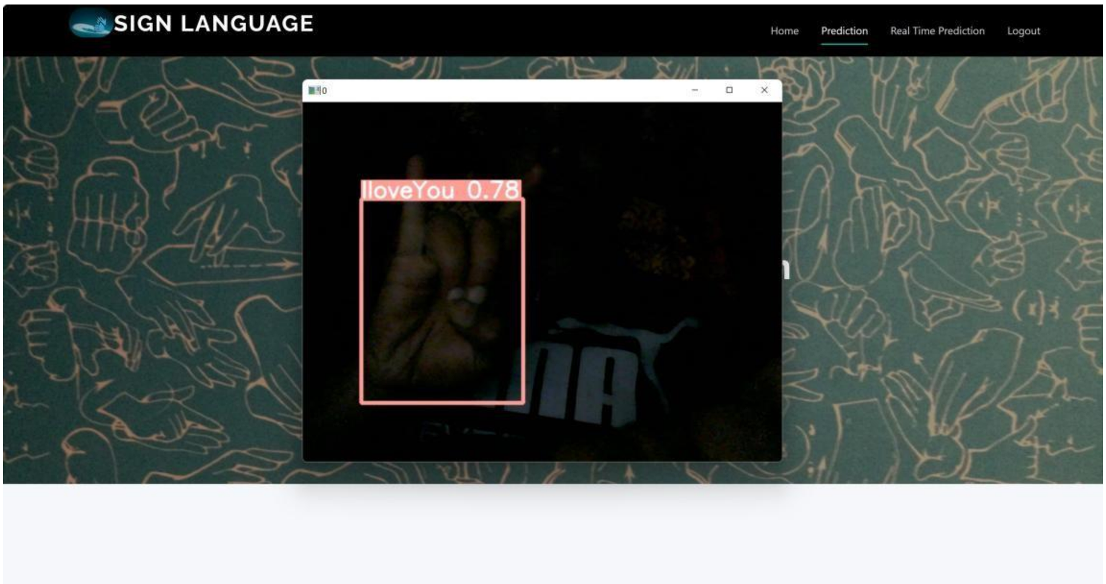

Figure 7: Real-time sign language gesture recognition using a webcam with the proposed system.

V. FUTURE SCOPE

Enhancements to the proposed sign language recognition system could include expansion of the recognition capabilities to include a larger vocabulary of both static and dynamic gestures in multiple sign languages, which would improve usability across a wider range of individuals. Future enhancements would be the incorporation of NLP (Natural Language Processing), which will enable the generation of full sentences from the recognised gestures and allow for more natural communication between people who use sign language. The system could also be optimised for use on mobile devices and embedded platforms, thereby enabling users to have access to real-time sign language recognition at any location. The addition of facial expressions and body posture in conjunction with the recognition of hand gestures would help to improve the accuracy of the recognition of complex signs. In addition, the addition of cloud-based learning and advanced deep learning models would support ongoing improvements in the performance of the system while also providing the capability for applications in education, healthcare, customer service and various forms of smart assistive technologies for people who have hearing and/or speech impairments.

VI. CONCLUSION

We have created a new, realistic application that can understand sign language and convert it to text and audio in real-time, through the use of machine learning and computer vision technologies. Not only will this application be able to interpret and understand sign language, but it can also seamlessly integrate with many different forms of technology and devices. This application works by identifying the position of a person's hands and displaying the image of their hands as a video through a viewfinder. We used open-source software and a webcam to collect the data needed for this application, which will help keep costs down. In short, predicting sign language goes beyond just being able to predict the hand gestures of a signer; it is also intended to create a more equitable, open society where everyone has an opportunity to communicate and prosper.

References

[1] Monika Chaturvedi, Bharti Kothari, Sandhya Singh, Vandana Rani Kujur, and Aastha Tiwari, "Sign Language Recognition," *International Journal for Research Trends and Innovation (IJRTI)*, 2022.

[2] Ahmed Sultan, Walied Makram, Mohammed Kayed, and Abdelmaged Amin Ali, "Sign Language Identification and Recognition," *De Gruyter*, 2022.

[3] Satwik Ram Kodandaram, N. Pavan Kumar, and Sunil G. L., "Sign Language Recognition," *ResearchGate*, 2021.

[4] Karan Bhavsar, Raj Ghatiya, Aarti Gohil, Devanshi Thakkar, and Bhumi Shah, "Sign Language Recognition," *International Journal of Research Publication and Reviews (IJRPR)*, 2021.

[5] Prof. Radha S. Shirbhate, Vedant D. Shinde, Sanam A. Metkari, Pooja U. Borkar, and Mayuri A. Khandge, "Sign Language Recognition Using Machine Learning Algorithm," *International Research Journal of Engineering and Technology (IRJET)*, 2020.

[6] Sulochana Nadgeri and Dr Arun Kumar, "Survey on Sign Language Recognition System," *Mahatma Education Society's Transactions and Journals' Conference Proceedings on CTFC*, ISBN, 2018.

[7] Omkar Jovalkar, Pratik Gaikar, and Pramod Gavali, "Sign Language Recognition Techniques: A Survey," *International Research Journal of Engineering and Technology (IRJET)*, 2020.

[8] Refat Khan Pathan, Munmun Biswas, Suraiya Yasmin, Mayeen Uddin Khandaker, Mohammad Salman, and Ahmed A. F. Youssef, "Sign Language Recognition Using the Fusion of Image and Hand Landmarks Through Multi-Headed Convolutional Neural Network," *Scientific Reports*, 2023.

[9] Ronglai Zuo, Fangugyun Wei, and Brian Mak, "Towards Online Sign Language Recognition and Translation," The Hong Kong University of Science and Technology and Microsoft Research Asia, 2024.