American Sign Language Real Time Detection Using TensorFlow and Keras in Python

Jayashre
Student
Department of Computer Science and Engineering
Shiv Nadar University
Chennai, India
jaya2004kra@gmail.com

***Abstract*—This paper presents a novel approach to enhance communication for individuals with hearing impairments. We propose a sign language detection program in Python that integrates image recognition techniques to synthesize American Sign Language (ASL). The program achieves accurate recognition by analysing and interpreting ASL gestures using a convolutional neural network. The detected gestures are mapped to textual representations and relayed to the display interface for seamless communication. Evaluation of a diverse data set demonstrates the system's robustness and effectiveness. The developed program, implemented in Python using OpenCV and TensorFlow, offers a scalable solution to improve accessibility and inclusivity for the hearing-impaired community.**

***Keywords*—ASL, Image Recognition, Keras, CNN, TensorFlow**

I. INTRODUCTION

Sign language is a visual form of communication that utilizes hand gestures and motions to convey meaning. It plays a vital role in enabling effective communication for individuals who are deaf or have hearing impairments, allowing them to express their thoughts, emotions, and interact with others in their community.

Inclusivity of individuals who are deaf and mute is crucial for building a society that values diversity and ensures equal opportunities. By promoting accessible communication methods, such as sign language interpretation and inclusive technologies, we can foster a more inclusive environment where the deaf and mute community can actively participate and contribute to all aspects of life.

Our approach to creating a sign language detection system consists of first creating a histogram (shown in Fig.6) and then reading the real time video input frame be frame and then comparing each frame to the dataset. After this the prediction is done using a model created by a CNN network whose accuracy is about 97%. Once the model predicts the sign, it is displayed to the user using the chatbot that also reads out the output.

The primary contributions of the work are as follows:

- ***ASL Detection Methodology***: Introduces an effective ASL gesture detection approach using threshold images and deep learning.
- ***Custom Deep Learning Model***: Develops a Keras-based deep learning model, improving ASL gesture recognition accuracy.
- ***Real-Time Recognition***: Enables real-time ASL recognition for immediate communication.
- ***Text-to-Speech Integration***: Integrates Pyttsx3 for converting ASL signs to spoken language.
- ***End-to-End Development***: Involves dataset creation, model development, and user-friendly GUI design depicted in Fig.1 and Fig.2.

This paper is structured as follows: Section 2 covers the framework's architecture, Section 3 discusses methodology, Section 4 focuses on visualization, Section 5 addresses data acquisition, Section 6 delves into analysis, Section 7 presents results, and Section 8 concludes with future work prospects. References are provided in Section 9.

II. LITERATURE REVIEW

In their study, S. Ikram & N. Dhanda et al. [9] demonstrated that deep learning, including convolutional neural networks, attains an impressive accuracy of 98% in detecting ASL alphabet symbols captured through a webcam. This research addresses communication challenges for the deaf-mute community and aims to expand its applications to encompass additional gestures and sentence formation. Likewise, in the research conducted by T. A. Siby et al. [10], the Real-Time Sign Language Recognition (RTSLG) system improves communication for individuals who have hearing and speaking disabilities. It achieves an 87% accuracy rate, offering text output for the deaf and audio output for the blind. Both studies highlight the importance of technology in bridging communication gaps, promoting accessibility, and reducing reliance on human interpreters during sign language interactions.

M. Deshpande et al. [5] adopt a comprehensive approach that integrates Media Pipe and LSTM deep learning techniques for the detection of sign language. This approach holds the potential to enhance communication for individuals who are deaf or hard of hearing. It encompasses

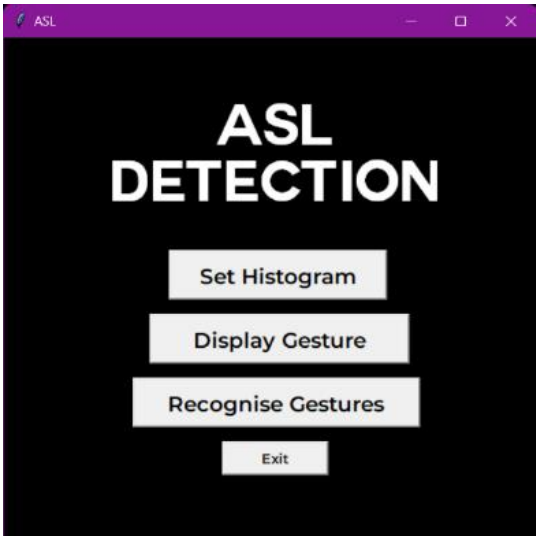

Fig.1 GUI Interface

Fig.2 The set of Gestures used to train our model.

various stages, including data processing, model input, and system implementation. A. Pardasani [1] has integrated American Sign Language (ASL) with chatbot technology, employing a combination of CNN and AIML. This integration provides precise communication support for individuals who are mute, particularly in the midst of prevailing voice recognition trends.

S. Almana & A. Al-Omary et al. [7] focus on the recognition of Arabic Sign Language (ArSL) to enhance communication between individuals who are hearing-impaired and those with normal hearing. Their approach encompasses both manual and non-manual sign elements, involving the use of Convolutional Neural Networks (CNNs) trained on the ArSL2018 dataset. Data preprocessing includes image conversion and ROI resizing. The CNN architecture achieves 95% accuracy, with various evaluation metrics presented. A. Tayal et al. [3] delve into the effects of the COVID-19 pandemic on education, with a particular focus on its impact on individuals with dyslexia. The A&N intelligent learning model promotes two-way student interaction and offers seven interactive educational projects using Brobject tracking technology. Technical implementation involves neural networks, Python, and OpenCV. Both papers show promise in enhancing communication and interactive learning in diverse educational contexts.

S. D. Boncolmo et al. [8] have developed a gender identification system that utilizes Convolutional Neural Networks on a Raspberry Pi. This system is designed for various applications, such as biometrics and surveillance. With a training accuracy of 96% and validation accuracy of 90%, it detects gender in real-time using approximately 2200 face images. The hardware setup includes a Raspberry Pi 4 Model B, camera module, and monitor, using two datasets for training and discussing limitations in ethnicity and age detection. A. S. Ghotkar [2] has developed a vision-based system for recognizing Indian Sign Language gestures, aimed at aiding the hearing-impaired. This system can accurately recognize gestures using a standard webcam. It consists of four modules, including hand tracking and feature extraction, offering real-time effectiveness and cost-effectiveness for improved communication and social integration among the deaf.

J. -H. Sun et al. [4] center their paper on the subject of hand gesture recognition within human-computer interaction systems. The paper covers various aspects, including segmentation, tracking, and recognition. It discusses various segmentation methods and proposes a fusion of skin color Gaussian mixture models and AdaBoost classifiers for improved accuracy. Camshift algorithm is chosen for real-time tracking, and a LeNet-5 neural network achieves 98.3% accuracy in recognition. The paper highlights the need to integrate techniques for better recognition and acknowledges challenges with complex gestures and 3D information limitations. Mampi Devi et al. [6] emphasize the importance of gesture recognition, particularly in the context of Manipuri Classical Dance. It underscores the importance of creating a dataset for this dance form and outlines the phases of gesture recognition. Various techniques like skeletonization, statistical methods, SVM, HMM, and FSM are discussed. The paper also showcases the broad applications of gesture recognition in fields such as computer vision, electronic interfaces, healthcare, and education, highlighting its practical importance.

III. METHODOLOGY

The ASL detection methodology (as shown in Fig.3) employed in this project involves comparing threshold images. The dataset consists of thresholder images representing each alphabet/number's ASL sign. The dataset can be either imported or recorded frame by frame, depending on the specific requirements. Once the dataset (shown in Fig.5) is prepared, it is divided into testing and training data subsets.

The training data is then processed and used to train a model that will be utilized for prediction. The model is built using Keras, a popular deep learning library. Keras provides a user-friendly interface for constructing neural networks and efficiently training them on large datasets.

To predict a particular gesture, frames are read from an input video stream, and these frames are converted into threshold images. The threshold images are then compared with the dataset using the trained model. This comparison helps in identifying the corresponding ASL sign for the input gesture.

IV. VISUALISATION

The process begins by capturing hand gestures using a live camera to create the dataset. The camera records images of individuals performing various ASL gestures. To enhance the quality of the captured images, preprocessing techniques such as Gaussian and median blur are applied. These techniques help to reduce noise and improve the clarity of the hand gestures in the images.

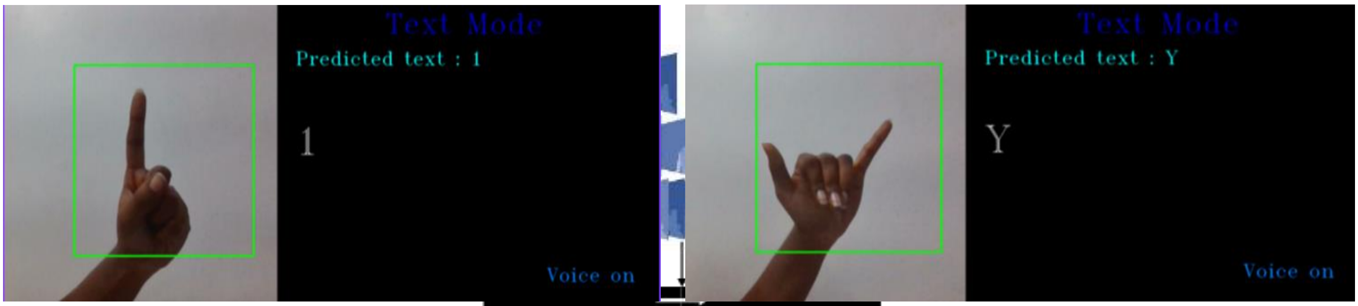

Fig. 4 S[redacted]splay

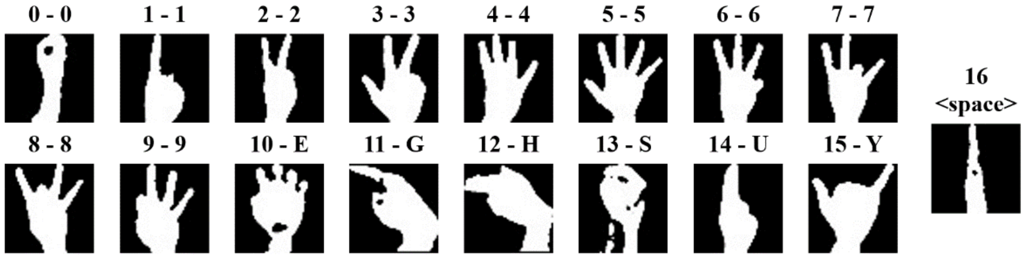

Fig.5 The gestures should be stored in the database along with their names.

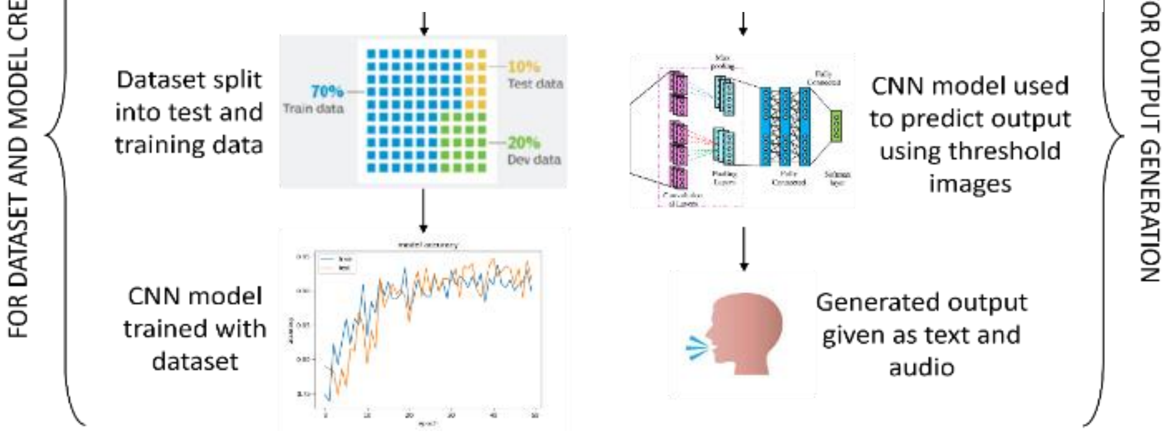

Fig.3 Architecture diagram

Training the model involves feeding the processed training data into the model and optimizing its parameters to learn patterns and features that are indicative of each ASL sign. The training process aims to minimize the model's prediction error and improve its accuracy in recognizing ASL gestures.

In this specific project, once the final output gesture is predicted, it is then spoken aloud using the Pyttsx3 Python library. Pyttsx3 provides a text-to-speech interface, allowing the system to convert the recognized ASL sign into spoken language, enhancing accessibility and understanding for individuals who may not be familiar with sign language. This is shown in Fig.4.

After applying blurring, thresholding is performed on the pre-processed images. This step converts the grayscale images into binary images, where the hand regions are highlighted. The thresholder image is then saved into the system, forming a part of the dataset.

To increase the diversity of the data’s et, the recorded images are flipped horizontally. Flipping the images creates mirrored versions of the gestures, effectively doubling the dataset's size. This augmentation technique helps to provide more varied examples of the hand gestures, allowing the model to learn from a wider range of perspectives.

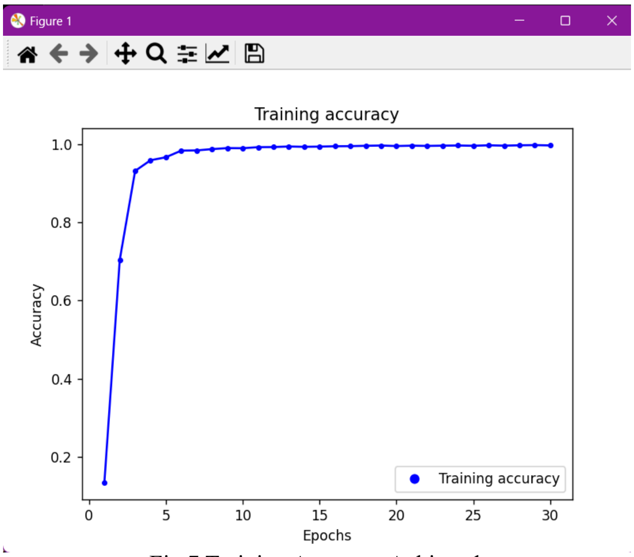

Fig.7 Training Accuracy Achieved

Once the dataset is prepared, it is loaded into the code environment. The dataset is then split into training and testing data. The training data is used to train a Keras model, which is a high-level neural networks API built on top of TensorFlow. Keras provides an intuitive interface for designing and training deep learning models, making it suitable for tasks like hand gesture recognition.

The Keras model, after being trained on the dataset, is imported into the final code for action recognition. The model's architecture and learned weights enable it to recognize and classify ASL gestures accurately. When a new input (e.g., a live video stream) is provided, the imported model can analyse the hand gestures in real-time and predict the corresponding ASL action or sign.

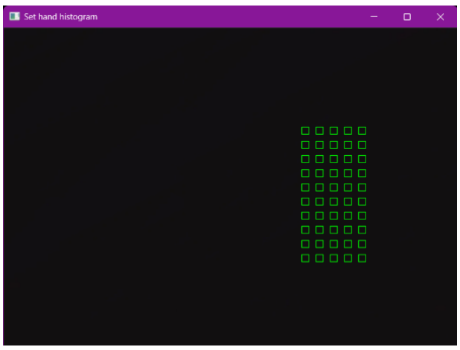

Fig.6 This is how to histogram is set.

Researchers and developers can utilize this data acquisition and modelling pipeline to build effective ASL gesture recognition systems. These systems hold promise in enhancing communication and accessibility for the deaf and hard-of-hearing community, bridging the gap between sign language and spoken language.

V. DATA AQUISATION

This section focuses on the data acquisition process, which plays a crucial role in capturing American Sign Language (ASL) gestures from proficient individuals. The process involves setting up the data collection environment, ensuring optimal lighting conditions, appropriate camera placement, and adhering to specific hand positioning guidelines.

To begin, creating a suitable data collection environment is essential. This includes ensuring adequate lighting conditions to enable clear visibility of the hand gestures. Proper lighting helps capture accurate hand shapes and movements, minimizing potential errors in the dataset. Additionally, the camera placement is crucial for capturing the gestures effectively. The camera should be positioned to have a clear view of the hands, ensuring that all hand movements are captured accurately.

During the recording process, a diverse range of ASL gestures should be considered to ensure the dataset encompasses various hand shapes, orientations, and movements. This diversity is important as ASL encompasses a rich vocabulary with numerous signs that vary in complexity and intricacy. By capturing a wide range of gestures, the resulting dataset becomes more comprehensive and representative of the full spectrum of ASL.

After acquiring the data, it can be used to create a framework for hand gesture recognition. This involves leveraging libraries such as TensorFlow and Keras, which provide powerful tools for training machine learning models. TensorFlow allows the creation of dataflow graphs, which describe how data moves through a computational graph. This capability is particularly useful for building models that process and recognize hand gestures based on the captured ASL data.

In addition to TensorFlow, scikit-learn is another valuable library that can be employed in the data analysis process. Scikit-learn offers a range of machine learning models for tasks such as regression, classification, and clustering. It also provides statistical tools to analyze these models, enabling researchers to gain insights into the performance and behavior of the trained models.

By utilizing these libraries and techniques, researchers can leverage the collected ASL dataset to train machine learning models that can accurately recognize and interpret ASL gestures. This advancement in technology holds the potential to bridge the communication gap between the deaf community and those who do not understand sign language, enhancing accessibility and inclusivity for all.

VI. ANALYSIS

In addition to the previous steps, the utilization of gesture analysis algorithms further enhances the system's capability to interpret and recognize ASL gestures accurately. These algorithms are specifically designed to analyze both the spatial and temporal characteristics of hand movements, extracting meaningful features and patterns that represent different ASL gestures.

By capturing information such as hand shape, orientation, movement trajectory, and timing, the gesture analysis algorithms enable the system to capture and comprehend the nuances of ASL gestures. These algorithms often leverage techniques such as image processing, computer vision, and machine learning to process the input data effectively.

During the training phase, the system is exposed to a comprehensive dataset of ASL gestures that cover a wide range of signs. This dataset serves as a reference, allowing the system to learn and identify the distinctive features associated with each gesture. Through this training process, the system becomes adept at recognizing and interpreting ASL gestures accurately. The accuracy for the same is shown in Fig.7.

By extracting relevant features from the input data and comparing them with the learned patterns from the training dataset, the system can make predictions and determine the corresponding ASL sign for a given gesture. The use of advanced machine learning techniques, such as deep learning models, can further enhance the accuracy and robustness of the gesture recognition system.

VII. RESULTS

The real-time American Sign Language (ASL) recognition system utilizing gesture analysis demonstrates impressive performance in accurately interpreting and recognizing ASL hand gestures. Extensive testing and evaluation of the system have been conducted to assess its effectiveness and reliability in real-world scenarios. To evaluate the system's performance, a diverse dataset comprising a wide range of ASL gestures was collected from proficient sign language users. The dataset encompassed various hand shapes, orientations, and movements, ensuring the system's ability to handle different sign variations. The user-friendly graphical user interface (GUI) significantly enhanced the usability and accessibility of the system. It was found that the GUI is intuitive and easy to navigate, enabling seamless interaction and control. The GUI facilitated various actions, including setting up the system, calibrating hand gestures, and accessing recognition results, making it user-friendly for individuals with different levels of technical expertise.

VIII. CONCLUSION

In conclusion, the real-time ASL recognition system utilizing computer vision and gesture analysis represents a significant advancement in enabling effective communication between ASL users and non-sign language users. By leveraging computer vision techniques, sophisticated gesture analysis algorithms, and a user-friendly GUI, the system demonstrates remarkable potential for improving accessibility, fostering inclusivity, and bridging the communication gap between diverse communities. Continued research and development in this field will pave the way for even more accurate and efficient ASL recognition systems in the future.

However, it's important to acknowledge some limitations. The system's accuracy heavily relies on a contrasting background and hand color, and when this contrast is lacking, the accuracy may decrease. To enhance accuracy in such scenarios, future developments could consider predicting sign language based on feature extraction or detection points from the hands rather than relying solely on threshold images.

Ultimately, the combination of gesture analysis algorithms and comprehensive training enables the system to facilitate seamless communication between ASL users and non-sign language users, empowering individuals proficient in ASL to express themselves naturally and enabling non-sign language users to understand and respond to ASL gestures effectively. This technological advancement contributes to fostering inclusivity and accessibility for the deaf and hard-of-hearing community, promoting equal opportunities for communication and understanding.

IX. REFERENCES

[1] A. Pardasani, A. K. Sharma, S. Banerjee, V. Garg and D. S. Roy, "Enhancing the Ability to Communicate by Synthesizing American Sign Language using Image Recognition in A Chatbot for Differently Abled," in *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*.

[2] A. S. Ghotkar, R. Khatal, S. Khupase, S. Asati and M. Hadap, "Hand gesture recognition for Indian Sign Language," in *2012 International Conference on Computer Communication and Informatics.*

[3] A. Tayal, "An Interactive Alphabet and Number learning system using OpenCV and CNN," in *2021 Asian Conference on Innovation in Technology (ASIANCON).*

[4] J. -H. Sun, T. -T. Ji, S. -B. Zhang, J. -K. Yang and G. -R. Ji, "Research on the Hand Gesture Recognition Based on Deep Learning," in *2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE).*

[5] M. Deshpande et al., "Sign Language Detection using LSTM Deep Learning Model and Media Pipe Holistic Approach," in *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*.

[6] Mampi Devi, Ananya Chakraborty, Alak Roy, and Dipanjoy Majumder, "Single-hand Gesture Recognition of Manipuri Classical Dance of India based on Skeletonization Technique," in *2023 International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC).*

[7] S. Almana and A. Al-Omary, "Real-time Arabic Sign Language Recognition using CNN and OpenCV," in *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT).*

[8] S. D. Boncolmo, E. V. Calaquian and M. V. C. Caya, "Gender Identification Using Keras Model Through Detection of Face," in *2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM).*

[9] S. Ikram and N. Dhanda, "American Sign Language Recognition using Convolutional Neural Network," in *2021 IEEE 4th International Conference on Computing, Power, and Communication Technologies (GUCON).*

[10] T. A. Siby, S. Pal, J. Arlina, and S. Nagaraju, "Gesture based Real-Time Sign Language Recognition System," in *2022 International Conference on Connected Systems & Intelligence (CSI)*.