Real Time Sign Language to Text Conversion

Project Report Submitted to

Shri Ramdeobaba College of Engineering &

Management, Nagpur in partial fulfillment of the award of

Degree of

Bachelor of Technology

In

Computer Science and Engineering

By

Mohammad Saify Sheikh (49)

Soham Bedi (66)

Himanshu Shrigiriwar (45)

Rugved Mhatre (60)

Guide

Prof. Shubhangi Tirpude

RCOEM

Shri Ramdeobaba College of
Engineering and Management, Nagpur

Department of Computer Science and Engineering

Shri Ramdeobaba College of Engineering and Management,

Nagpur 440013

(An Autonomous Institute affiliated to Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur)

April 2023 - 24

Real Time Sign Language to Text Conversion

Project Report Submitted to

Shri Ramdeobaba College of Engineering &

Management, Nagpur in partial fulfillment of the award of

Degree of

Bachelor of Technology

In

Computer Science and Engineering

By

Mohammad Saify Sheikh (49)

Soham Bedi (66)

Himanshu Shrigiriwar (45)

Rugved Mhatre (60)

Guide

Prof. Shubhangi Tirpude

Department of Computer Science and Engineering

Shri Ramdeobaba College of Engineering and Management,

Nagpur 440013

(An Autonomous Institute affiliated to Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur)

April 2023 - 24

Shri Ramdeobaba College of Engineering and Management, Nagpur

(An Autonomous Institute affiliated to Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur) Department of Computer Science and Engineering

CERTIFICATE

This is to certify that the Project Report on " **Real Time Sign Language to Text Conversion** " is a bonafide work of Mohammad Saify Sheikh, Himanshu Shrigiriwar, Soham Bedi, Rugved Mhatre submitted to the Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur in partial fulfillment of the award of a Degree of Bachelor of Engineering. It has been carried out at the Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur during the academic year 2023-2024.

Date :

Place : Nagpur

Prof. Shubhangi Tirpude	Dr. Ramchand Hablani
(Project Guide)	(HOD)
Department of Computer Science and Engineering	Department of Computer Science and Engineering

Dr. R. S.Pande

(Principal)

DECLARATION

I, hereby declare the project report “**Real Time Sign Language to Text Conversion**.” submitted herein, has been carried out in the Department of Computer Science and Engineering of Shri Ramdeobaba College of Engineering & Management, Nagpur. The work is original and has not been submitted earlier as a whole or part for the award of any degree / diploma at this or any other Institute / University.

Date :

Place : Nagpur

Name of Student	**Roll Number**	**Signature**
Mohammad Saify Sheikh	49	
Soham Bedi	66	
Himanshu Shrigiriwar	45	
Rugved Mhatre	60	

APPROVAL SHEET

This report entitled

Real Time Sign Language to Text Conversion

By

Mohammad Saify Sheikh

Soham Bedi

Himanshu Shrigiriwar

Rugved Mhatre

is approved for the degree of Bachelor of Engineering.

Name & Signature of Supervisor(s)

Name & Signature of External Examiner(s)

Name & Signature of HOD

Date:

Place: Nagpur

ACKNOWLEDGEMENT

The project is a combined effort of a group of individuals who synergize to contribute towards the desired objectives. Apart from the efforts by us, the success of the project shares an equal proportion on the engagement and guidance of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project. We would like to extend my heartfelt thanks to Prof. Shubhangi Tirpude for her support and encouragement during this project. Her knowledge, expertise, and willingness to help have been instrumental in the successful completion of this work. We would like to show our greatest appreciation towards Dr. Ramchand Hablani, Head, Department of Computer Science and Engineering, RCOEM, Nagpur for providing us with the facilities for completing this project and for his constant guidance. We would like to express our deepest gratitude to Dr. Rajesh Pande, Principal, RCOEM, Nagpur for providing us the opportunity to embark on this project. Finally, we extend our gratitude to all the faculty members of the CSE department who have always been so supportive and providing resources needed in our project development. We are very grateful to all of them who have unconditionally supported us throughout the project.

Date: — Projectees

ABSTRACT

In a world where gestures speak volumes and silence carries meaning, we introduce a groundbreaking innovation: a blend of technology and compassion that translates sign language into written text, and then to audio in real-time. Welcome to a realm where hands shape words, and inclusivity finds its voice. This report sheds light on the journey of creating a transformative solution, set to bridge communication gaps and champion accessibility for everyone.

Our system is skilled at recognizing 26 letters, operating instantly to capture and interpret hand gestures accurately. By using advanced machine learning and computer vision, it identifies characters and forms words on-the-spot, allowing for complete sentences to be constructed effortlessly.

This report not only explains the technical details of our solution but also highlights its profound impact on society. By breaking communication barriers for those with hearing impairments, our system promotes inclusivity, ensures equal access to information, and empowers users to participate fully in various social and professional contexts.

The development of this real-time sign-to-text conversion system represents a significant step towards creating an inclusive digital space. Its smart design, smooth operation, and societal benefits make it a symbol of progress, ready to redefine communication norms and enhance accessibility for all.

Keyword: Gesture Recognition, Text Conversion, Machine Learning, Convolutional Neural Network.

TABLE OF CONTENT

Contents **Page No.**

LIST OF FIGURE

LIST OF ABBREVIATION

Abbreviation	Expansion
CNN	Convolutional Neural Network
CV	Computer Vision
GUI	Graphical User Interface

CHAPTER 1

INTRODUCTION

Sign language is a rich and expressive form of communication used by the deaf and hard of hearing community worldwide. However, the accessibility of sign language can be limited due to the lack of widespread understanding and interpretation. Sign language to text conversion with audio output is a groundbreaking technology that aims to bridge this communication gap by providing a means to translate sign language gestures into written text and spoken words in real-time.

This innovative technology holds immense promise in enhancing communication between individuals who use sign language and those who do not. By converting sign language into text, it enables deaf individuals to effectively communicate with a broader audience, including those who may not be proficient in sign language. Moreover, the inclusion of audio output adds an additional layer of accessibility by providing spoken translations, making the communication process more seamless and inclusive.

1.1 Problem Statement:

Creating an advanced, real-time sign language to text conversion system represents a significant technological challenge with profound societal implications. This ambitious project seeks to address the pressing need for effective communication solutions for individuals with hearing impairments, aiming to bridge the gap between sign language users and those who rely on written or spoken language. By developing a robust and innovative solution, this endeavor aims to empower individuals with hearing impairments to communicate effortlessly and participate fully in various aspects of society, from education and employment to social interactions and beyond. The overarching goal is to foster inclusivity, promote accessibility, and enhance the quality of life for individuals with hearing impairments by enabling seamless and accurate translation of sign language gestures into written text in real-time.

1.2 Motivation:

The motivation behind real-time sign language to text conversion lies in addressing the Significant challenges and limitations present in current methods of interpreting and transcribing sign language. Individuals who rely on sign language communication face barriers in effectively expressing themselves and understanding others in various contexts, including education, healthcare, and public services. Traditional methods of transcription may not always capture the intricacies of sign language, leading to misinterpretations and misunderstandings.

By leveraging advanced technologies such as machine learning algorithms and computer vision techniques, we aim to establish a more reliable and efficient system for real-time sign language to text conversion. The adoption of real-time sign language to text conversion offers numerous benefits, including enhancing accessibility for individuals with hearing impairments, facilitating seamless communication across linguistic communities, and promoting inclusivity in various settings. By harnessing the power of technology, we can empower sign language users to communicate effectively and participate fully in society.

In conclusion, the motivation for exploring real-time sign language to text conversion arises from the imperative to address the challenges and limitations in current communication methods. Through innovation and collaboration, we can create a more inclusive and accessible world where communication barriers are overcome, and everyone has equal opportunities to engage and connect.

1.3 Overview:

Our project focuses on developing a real-time sign language to text as well as audio conversion system, catering to the diverse communication needs of individuals who rely on sign language. This innovative system aims to bridge communication gaps by converting sign language gestures into both written text and spoken words, thereby enhancing accessibility, promoting inclusivity, and facilitating seamless communication across diverse linguistic communities.

The real-time sign language to text and audio conversion system leverages advanced machine learning algorithms, computer vision techniques, and audio processing technologies to accurately interpret and transcribe sign language gestures into written text while simultaneously generating clear and intelligible audio output. This dual functionality enables users to choose their preferred mode of communication, whether through written text or spoken words, depending on their individual preferences and requirements.

Furthermore, the system is designed to be adaptable and customizable, allowing users to tailor the output to their specific needs and preferences. This includes features such as adjustable text font and size, customizable vocabulary, adjustable audio volume, voice gender selection, and support for different language accents, ensuring that the system can accommodate diverse users and communication contexts.

Our project aims to enhance accessibility and promote inclusivity for individuals with hearing impairments by providing accurate and clear output for sign language gestures,both in written text and audio format. Through innovation, collaboration, and user-centric design, we aspire to create a comprehensive solution that empowers sign language users to communicate effectively and participate fully in society, ultimately contributing to a more inclusive and accessible world.

1.4 Objectives

Following are objective for our project:

- Accuracy and Efficiency: Our first objective is to ensure that the real-time sign language to text conversion system achieves high levels of accuracy and efficiency. By leveraging advanced machine learning algorithms and computer vision techniques, we aim to accurately recognize and interpret sign language gestures, facial expressions, and body movements, minimizing errors and misunderstandings in the transcription process.
- Real-Time Capability: Another key objective is to develop the system with real- time capabilities, enabling instantaneous conversion of sign language gestures into text. This real-time functionality is essential for facilitating seamless communication in various settings, including educational, healthcare,and public service environments, where timely and efficient communication is critical.
- Seamless Audio Output: We aim to ensure that the audio output generated by the system is seamless and natural-sounding, enhancing the overall user experience and facilitating effective communication. This includes optimizing the audio processing algorithms to produce high-quality audio output that accurately reflects the intended meaning of the sign language gestures.
- User-Friendly Interface: Finally, our objective is to develop a user-friendly interface that is intuitive and easy to use, ensuring that the system is accessible to individuals with varying levels of technical proficiency. This includes designing clear and intuitive visual feedback for users, providing easily navigable menus and controls, and incorporating accessibility features such as screen reader compatibility.
- Overall, the objective of our project is to develop a real-time sign language to text conversion system that addresses the unique needs and challenges of sign language communication, ultimately promoting accessibility, inclusivity, and equal opportunities for all individuals. Through innovation, collaboration, and user- centric design, we aim to create a solution that empowers sign language users to communicate effectively and participate fully in society.

1.5 Applications

- Education: In educational settings, our system can facilitate seamless communication between students who use sign language and teachers or peers who rely on text-based communication. It can be integrated into classrooms, lecture halls, and online learning platforms to ensure equal access to educational resources and opportunities for all students, regardless of their communication preferences.
- Healthcare: In healthcare facilities, our system can improve communication between healthcare providers and patients who use sign language. It can be used during medical consultations, examinations, and treatment sessions to ensure accurate transmission of information, enhance patient understanding, and promote patient-centered care.
- Public Services: Our system can also be utilized in public service settings, such as government offices, libraries, and transportation hubs, to facilitate communication with individuals who use sign language. It can help government agencies and service providers deliver accessible and inclusive services to all citizens, fostering a more equitable and inclusive society.
- Customer Service: In businesses and commercial establishments, our system can enhance customer service experiences for individuals who use sign language. It can be integrated into customer support channels, such as helplines,websites, and mobile applications, to ensure that all customers receive timely and accurate assistance, regardless of their communication preferences.
- Emergency Response: Our system can play a crucial role in emergency situations, where clear and effective communication is essential for ensuring public safety and providing assistance to individuals with hearing impairments. It can be integrated into emergency response systems, such as 911 call centers and disaster relief operations, to facilitate communication between emergency responders and individuals who use sign language.

CHAPTER 2
LITERATURE REVIEW

S, Thakur at el [1] proposed a deep learning model trained on the ASL i.e the American Sign Language which will take action in the form of American Sign Language as input and translate it into text. Convolution Neural Network based VGG16 architecture is used as well as a TensorFlow model for image classification. There has been an improvement in accuracy from 94% of CNN to 98.7% by Transfer Learning.

T. Kemkar at el [2] proposed a research based in ASL fingerprint-based real-time method utilizing Convolutional Neural Networks (CNN). The approach involves passing the hand through a filter first, and then a classifier that determines the type of hand gesture once the filter has been applied. It is implemented using CNN as it is very effective at addressing computer vision issues. This proposed model achieved 99% accuracy on the first (MNIST) dataset and 96% on the second (ASL) dataset.

Kavya Dharshini at el [3] suggested a method to recognize the poses and hand gestures of all kinds of sign languages available worldwide and convert them into text and audio and making the model can be further improved by adding various language conversions for better performance. system bilingual. In this research work, the conversion of sign language to text and audio is carried out using techniques such as Media Pipe Holistic, Drawing Landmarks, Open CV, LSTM Neural Network, Google Translator, and GTTS in order to achieve a good accuracy of 98%. This Abey Abraham at el [4] makes use of an Arduino Uno board, a few flex sensors and an Android application to enable effective communication amongst the users. Using the flex sensors, gestures made by the wearer is detected and then according to various pre-defined conditions for the numerous values generated by the flex sensors, corresponding messages are sent using a Global System for Mobile (GSM) module to the wearer's android device, which houses the application that has been designed to convert text messages into speech. The GSM module is also used to send the sensor inputs to a cloud server and these values are taken as input parameters into the neural network for a time series-based prediction of gestures. The system is designed to continually learn devices and improve reliability by monitoring every individual's behavior at all times.

A. Adeyanju at el [5] analyzes 649 publications related to decision support and intelligent systems on sign language recognition (SLR) from the Scopus database. Moreover, reviews of techniques for vision-based sign language recognition are presented. Various features extraction and classification techniques used in SLR to achieve good results are discussed. The literature review presented in this paper shows the importance of incorporating intelligent solutions into the sign language recognition systems and reveals that perfect intelligent systems for sign language recognition are still an open problem. Overall, it is expected that this study will facilitate knowledge accumulation and creation of intelligent-based SLR and provide readers, researchers, and practitioners a roadmap to guide future direction.

Bagane at el [6] have presented the technology which accurately recognizes and instantly translates ASL signals into written text using cutting-edge computer vision and machine learning algorithms. The system is divided into the letter recognition model, a gesture recognition module, and a text generating module. Then, using the recognized movements, the text production module produces text. Their study describes how ASL to text converters might be used in accessibility services, education, and ordinary communication.

Akano at el [7] suggested the techniques of image segmentation and feature detection for the implementation. The system goes through various phases such as data capturing using KINECT sensor, image segmentation, feature detection and extraction from ROI, supervised and unsupervised classification of images with K-Nearest Neighbor (KNN)-algorithms and text-to-speech (TTS) conversion. The combination of FAST and SURF with a KNN of 10 also showed that unsupervised learning classification could determine the best matched feature from the existing database. The system achieved a 78% accuracy of unsupervised feature learning. The success of this work can be attributed to the effective classification that has improved the unsupervised feature learning of different images. The pre-determination of the ROI of each image using SURF and FAST, has demonstrated the ability of the proposed algorithm to limit image modeling to relevant regions within the image.

Zahid H at el collected a dataset of Urdu sign language (USL) and tested it through a machine learning classifier. The USL dataset which comprises 1,560 images was created by photographing various hand positions using a camera. This work provides a strategy for automated identification of USL numbers based on a bag-of-words (BoW) paradigm. For classification purposes, support vector machine (SVM), Random Forest, and K-nearest neighbor (K-NN) are used with the BoW histogram bin frequencies as characteristics. The proposed technique outperforms others in number classification, attaining the accuracies of 88%, 90%, and 84% for the random forest, SVM, and K-NN respectively.

	Method	Accuracy	Dataset used	Conclusion	Future Scope
Sign Language to Text Conversion in Real Time using Transfer Learning [1]	CNN model and VGG16 model	98.7% by Transfer Learning an improvement of 5%	ASL (American Sign Language)	Utilizing VGG16 the model achieved superior performance.	Diversify the model by training over the Indian sign language dataset, British sign language dataset, etc. For better result, augmentation techniques can be used.
Sign Language to Text Conversion using Hand Gesture Recognition [2]	CNN	96 %	MNIST data set	categorize hand gestures with a high degree of accuracy.	To increase accuracy & variety of dataset

Conversion of Sign Language to Text and Audio Using Deep Learning Techniques [3]	Media Pipe Holistic, Drawing Landmarks, Open CV, LSTM Neural Network, Google Translator, and GTTS	98%	American language dataset	recognize the poses and hand gestures of all kinds of sign languages available worldwide and convert them into text and audio and making the system bilingual.	to add various language conversions for better performance.
Machine learning methods for sign language recognition: A critical review and analysis [5]	Histogram Equalizatio n (HE), Adaptive Histogram Equalizatio n (AHE), Logarithmic Transformatio n	96.96	America n language dataset	Different research has been done on words, alphabets & numbers	More research in future is required for sentence recognition in sign language

An innovative ASL-to-Text Conversion System Leveraging Computer Vision and Machine Learning for Enhanced Communication [6]	CNN, ResNet 50	98.55	American language dataset		Develop the adaptive features to make the application functional in diverse environments, including low-light conditions or varying camera angles, to increase usability in different settings.
Deep Learning in Sign Language Recognition: A Hybrid Approach for the Recognition of Static and Dynamic Signs	YOLOv6, LSTM, CNN	92%	American language dataset	classification of different signs according to sign language and shows remarkable accuracy in real time.	To combine the developed models with fuzzy logic to deal with uncertainty in the outputs, to improve the detection rate in the presence

					of noise and changes in lighting and contrast.
A computer vision-based system for recognition and classification of Urdu sign language dataset[9]	SVM, Random Forest ,KNN	88%, 90%, and 84% for the random forest, SVM, and K-NN respectively.	Numerous orientations and forms photographs (hand configuration) are taken using a mobile camera.	USL number identification technique was constructed by extracting characteristics from histograms of bag-of-words (BoW)	To add more images and make it more diversified by adding USL alphabets.

CHAPTER 3

METHODOLOGY

3.1 Proposed Methodology

The proposed methodology begins with the collection of a diverse dataset comprising various sign language gestures, covering alphabets and common phrases. Following data collection, preprocessing steps are employed to prepare the dataset for model training. This involves resizing and normalizing the images to enhance the model's ability to generalize. Subsequently, Convolutional Neural Networks (CNNs) are employed for model training, leveraging their capability to effectively recognize and interpret sign language gestures. Once the model is trained, focus shifts to the development of a user-friendly Graphical User Interface (GUI). The GUI facilitates seamless interaction with the trained model, enabling real-time sign language interpretation. Finally, the system undergoes rigorous testing to validate its performance before deployment, ensuring that it effectively promotes inclusivity and accessibility for users.

3.2 Data Collection

At the beginning of our project, we focused on making a big collection of different hand gesture pictures. These pictures were essential for teaching our sign language recognition system. We used a special computer program written in Python to gather a wide range of images showing all the letters of the English alphabet and some common phrases.

3.2.1 Dataset Creation:

We carefully collected lots of pictures showing hand gestures, making sure we didn't miss anything. Using our special Python program, we made it easier to take pictures of gestures for every letter. This thorough process covered all sorts of hand positions and angles, making our dataset rich and varied. It helped us train our model better.

3.2.2 Preprocessing Techniques:

Employing cutting-edge computer vision techniques, we subjected the collected images to a series of preprocessing steps aimed at enhancing their quality and consistency. Techniques such as grayscale conversion and resizing were judiciously applied to standardize the images, thereby ensuring uniformity across the dataset. By mitigating variability and harmonizing image characteristics, we laid the groundwork for optimal model performance.

3.2.3 Dataset Partitioning:

Understanding how vital it is to check how well our model works, we split our collection of pictures carefully into two parts: one for training and the other for testing. We used 80% of the images to teach the model and kept the other 20% aside to see how good it was. This way, we made sure the model learned well and could handle different situations effectively.

Fig. 1.1: Dataset

3.3 Modeling Approach

Using our carefully collected dataset, we built a special computer program called a Convolutional Neural Network (CNN) for recognizing sign language. We used a powerful tool called Keras to design this program, making sure it could understand even the most complex hand movements and match them with the right letters and numbers accurately and quickly

3.3.1 Convolutional Layers:

At the heart of our CNN model lie multiple convolutional layers, each meticulously designed to extract and capture spatial features from input hand gesture images. Convolutions are the cornerstone of CNNs, performing localized operations to detect patterns and features within the input data. Mathematically, a convolution operation can be expressed as:

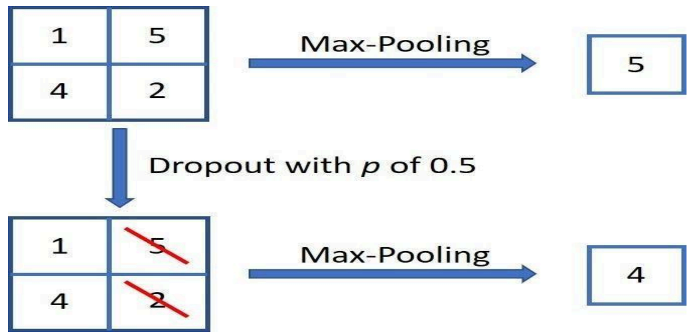
$$Output = Convolution(Input,Kernel)+Bias$$

Here, the kernel represents a learnable parameter matrix applied to the input data, and the bias term introduces a translation invariant component to the output. By cascading multiple convolutional layers, our model can hierarchically extract increasingly abstract and complex features, enabling robust pattern recognition and classification.

3.3.2 Max-Pooling and Dropout Layers:

Following each convolutional layer, we incorporated max-pooling and dropout layers to further enhance feature extraction and mitigate overfitting. Max-pooling layers downsample the feature maps, retaining only the most salient information while discarding redundant details. This process aids in reducing computational complexity and increasing translational invariance, thereby improving the model's robustness to input variations.

Additionally, dropout layers were strategically inserted to regularize the model and prevent overfitting. Dropout randomly deactivates a fraction of neurons during training, forcing the network to learn more robust and generalizable features. By selectively dropping neurons, dropout regularization encourages the model to avoid reliance on specific neurons, thus enhancing its ability to generalize to unseen data



1.2 : Max Pooling

3.3.3 Fully Connected Layers:

Towards the end of the network architecture, we incorporated fully connected layers to facilitate high-level abstraction and pattern recognition. These dense layers integrate the extracted features from the convolutional layers, enabling the model to capture complex relationships and correlations within the data. Mathematically, the output of a fully connected layer can be represented as:

$$\text{Output} = \text{Activation}(\text{Weights} \times \text{Input} + \text{Bias})$$

Here, the weights represent learnable parameters that are optimized during training to minimize the model's loss function. By applying non-linear activation functions such as ReLU (Rectified Linear Unit) or sigmoid, the fully connected layers introduce non-linearity into the model, enabling it to learn complex decision boundaries and make accurate predictions.

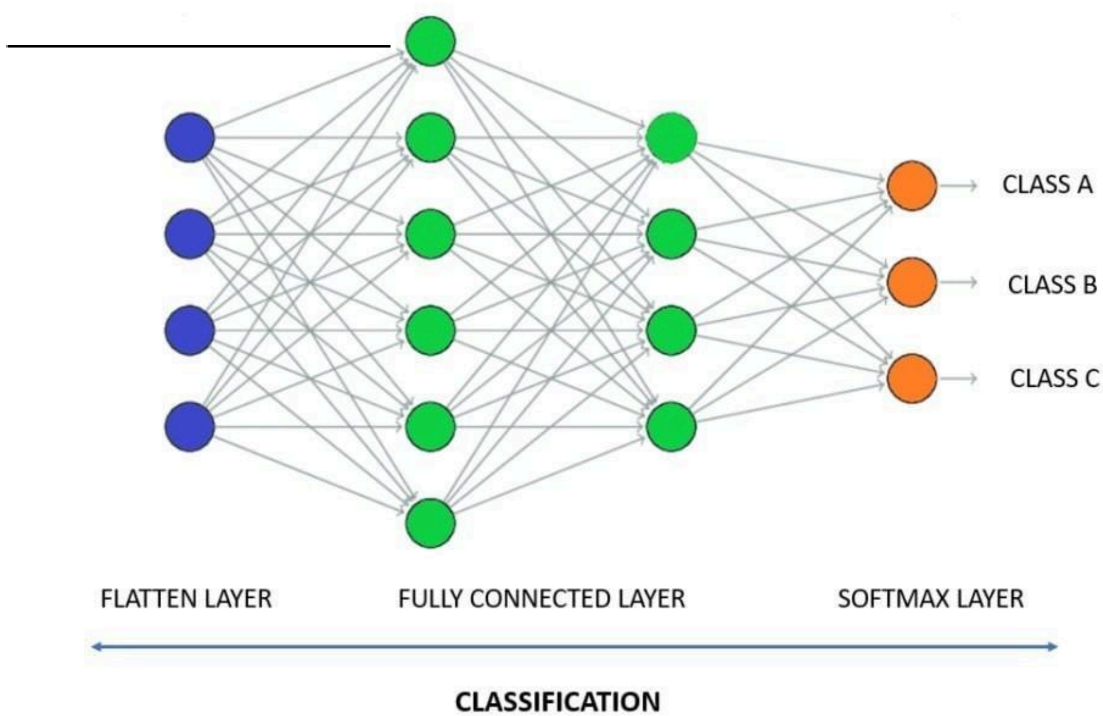


Fig 1.3: Fully Connected Layer

Output Layer:

At the culmination of our CNN architecture lies the output layer, where the final classification decisions are made. Leveraging the softmax activation function, the output layer assigns probabilities to each class, indicating the likelihood of the input hand gesture belonging to a particular character category. The softmax function computes the probability distribution across multiple classes, ensuring that the sum of probabilities equals one. Mathematically, the softmax function can be expressed as:

$$P(class_i) = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}}$$

In summary, our meticulously crafted CNN architecture, enriched with convolutional, max-pooling, dropout, and fully connected layers, embodies a sophisticated framework for sign language recognition. Through a synergy of mathematical formalism, computational efficiency, and algorithmic ingenuity, our model stands poised to revolutionize real-time communication for individuals with diverse communication needs.

3.4 Real-Time Inference

At the core of our project lies the real-time inference system, a user-friendly interface designed to seamlessly convert sign language into text. This intuitive application leverages Python's tkinter library for GUI construction and OpenCV for video processing.Upon launching the application, users are greeted with a live video feed capturing their hand gestures. The interface features:

Letter Row: Displays the currently recognized letter based on hand gestures.

Word Row: Shows the ongoing formation of words as gestures are made.

Sentence Row: Displays the evolving sentence constructed from successive gestures.

Audio Button: Converts the interpreted text into speech for enhanced accessibility.

The system swiftly analyzes hand movements, recognizes characters, and updates the displayed letters, words, and sentences in real-time. This seamless interaction fosters natural and fluid communication, promoting inclusivity and accessibility.

3.5 Technology Stack

Python: The cornerstone of our project, Python provides a versatile and dynamic programming environment.

Keras with TensorFlow Backend: Leveraging Keras, a high-level neural networks API, along with TensorFlow, our project constructs and trains convolutional neural network (CNN) models for sign language recognition.

Tkinter: Tkinter, Python's native GUI toolkit, enables the development of an intuitive and interactive user interface for real-time sign language interpretation.

Scikit-learn: a versatile machine learning library, complements our modeling pipeline with additional algorithms and utilities.

TensorFlow: TensorFlow is an open-source machine learning framework developed by Google for building and training neural networks.

Pyttsx3: Works offline, supports multiple TTS engines (including SAPI5, NSSpeechSynthesizer, and eSpeak), easy to use.

Flask : Flask is a lightweight web application framework in Python used for building web applications.

MongoDB : MongoDB is a NoSQL database known for its flexibility, scalability, and ability to store data in a JSON-like format.

CHAPTER 4

IMPLEMENTATION

The implementation phase of the Real-time Sign Language to Text Conversion System involved translating the conceptual design into a practical solution. This chapter outlines the steps taken to execute the plan, configure necessary infrastructure, integrate software components, and conduct rigorous testing to ensure the system meets the desired specifications.

4.1 Data Collection and Preprocessing:

Before training the Convolutional Neural Network (CNN) model, a dataset of sign language gestures was collected using a Python script named collect.py**.** This script utilized OpenCV to capture images of hand gestures through the camera interface. Each captured image was labeled and stored in a subfolder, representing a specific sign language letter. Data preprocessing techniques such as resizing, normalization, and augmentation were applied to enhance the quality and diversity of the dataset.

4.2 Model Training:

The collected dataset was used to train the CNN model for gesture recognition. TensorFlow with Keras was employed to design and train the model architecture. The CNN model consisted of multiple convolutional and pooling layers, followed by fully connected layers for classification. Various hyperparameters such as learning rate, batch size, and number of epochs were fine-tuned to optimize the model's performance. The training process involved splitting the dataset into training and validation sets to monitor the model's accuracy and prevent overfitting.

4.3 Real-time Sign Language Detection:

To enable real-time sign language detection, we first trained a deep learning model using Keras. The model was trained on a dataset of sign language images, with each image labeled with the corresponding sign language gesture. After training the model, we saved its architecture and weights to JSON and H5 files, respectively.

During real-time sign language detection, the trained model was loaded from the saved files. OpenCV was used to capture video from the webcam, and each frame was processed to extract features. The trained model then made predictions on these features to identify sign language gestures in real-time. The predicted gestures were displayed on a Tkinter GUI interface, providing a seamless user experience.

4.4 Gesture Recognition and Text Prediction:

Once the hand region is detected, the trained CNN model is applied to recognize the gesture being made by the hand. The model predicts the corresponding letter of the sign language alphabet. The recognized letter is then appended to a queue to form words and sentences. Additionally, a blank symbol is introduced to delineate between words and sentences, enhancing the readability of the converted text.

4.5 Suggestion Feature :

In addition to its core functionalities, our system features a suggestion feature aimed at enhancing user experience and improving communication efficiency. Leveraging MongoDB as our database solution, we store a comprehensive list of letters within our repository. When a user inputs a word, our system dynamically queries the MongoDB database to identify potential matches with the existing letter list. Upon finding matches, the system intelligently suggests relevant words or phrases based on the user's input. This feature not only facilitates faster typing but also assists users in forming coherent sentences, thereby streamlining the communication process. Through the seamless integration of MongoDB and intelligent suggestion algorithms, our system empowers users to express themselves more fluidly and effectively in sign language.

4.6 Graphical User Interface (GUI) Implementation:

A GUI was developed using Tkinter to provide a user-friendly interface for the system. The GUI displays the real-time video feed from the camera, the detected hand region, recognized letters, and formed words/sentences. Users can interact with the system through the GUI, observing the conversion of sign language gestures into text in real-time.

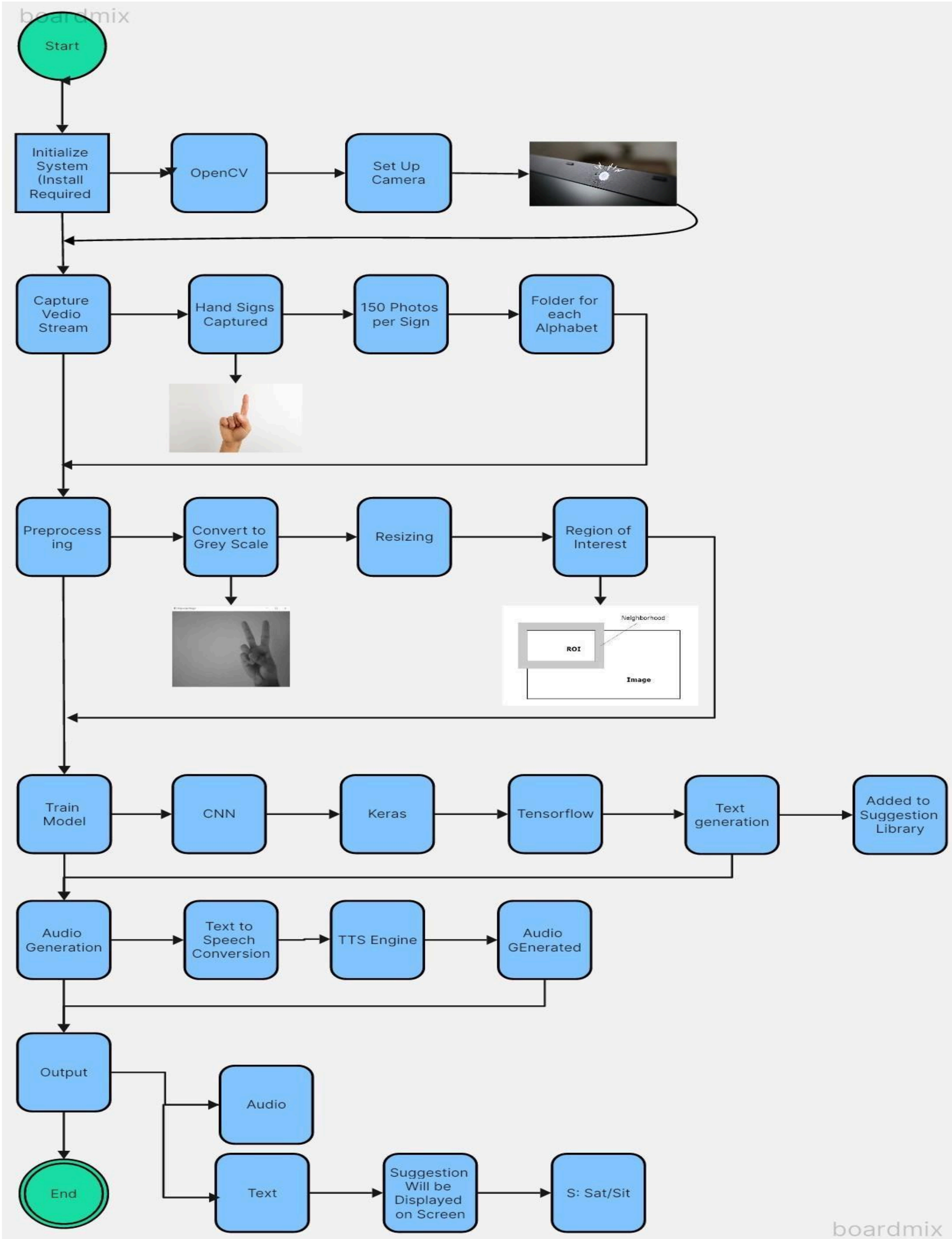


Fig 1.4: Proposed Working Flowchart

CHAPTER 5

RESULTS

In this section, we present a comprehensive analysis of the performance and efficacy of our real- time sign-to-text conversion system. Leveraging a meticulously crafted Convolutional Neural Network (CNN) architecture and a bespoke graphical user interface (GUI), our system embodies a culmination of cutting-edge technology and innovative design. Through a series of experiments and evaluations, we delve into the quantitative and qualitative aspects of our solution, shedding light on its robustness, accuracy, and real-world applicability.

5.1 Training Dynamics:

The training dynamics of our CNN model were analyzed using TensorBoard, a powerful visualization tool that offers unparalleled insights into the training process. Through dynamic graphs and visualizations, we scrutinized key training parameters such as loss, accuracy, and learning rate, unraveling the intricate nuances of the model's convergence behavior and optimization trajectory.

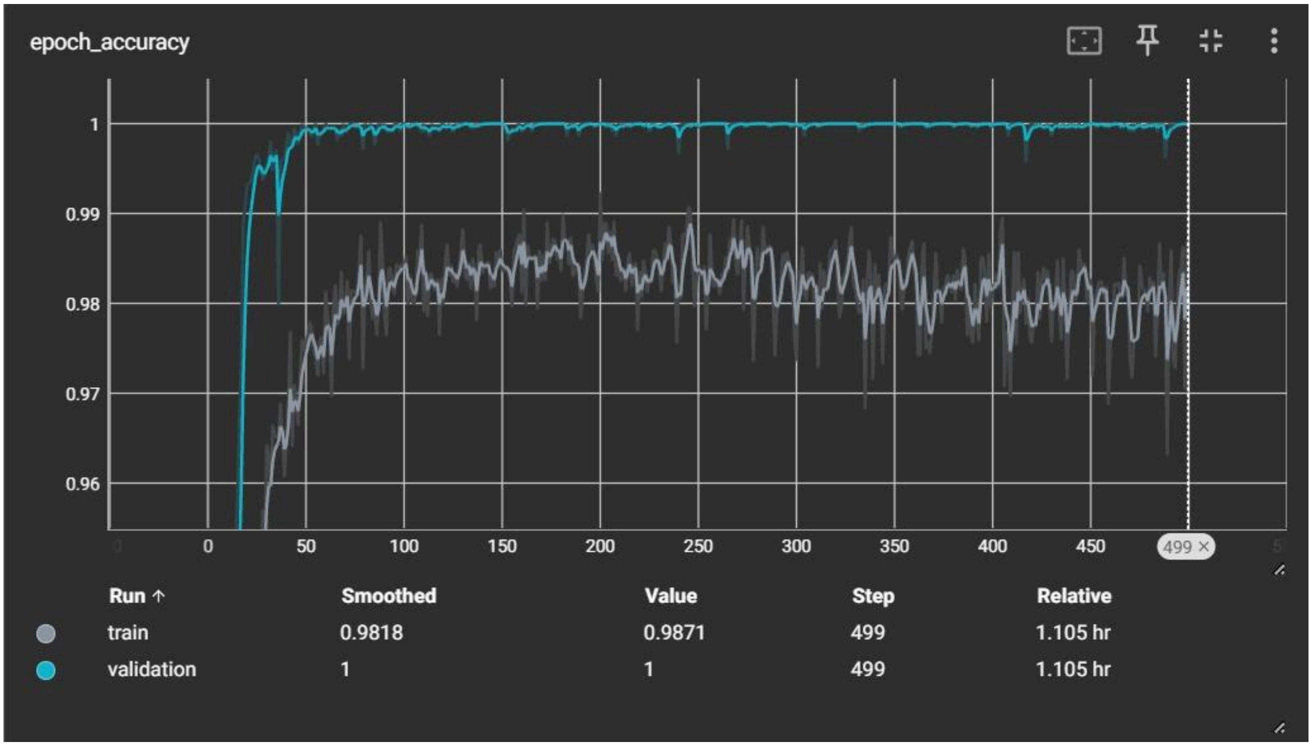


Fig. 1.5:Accuracy Vs Epoch Graph

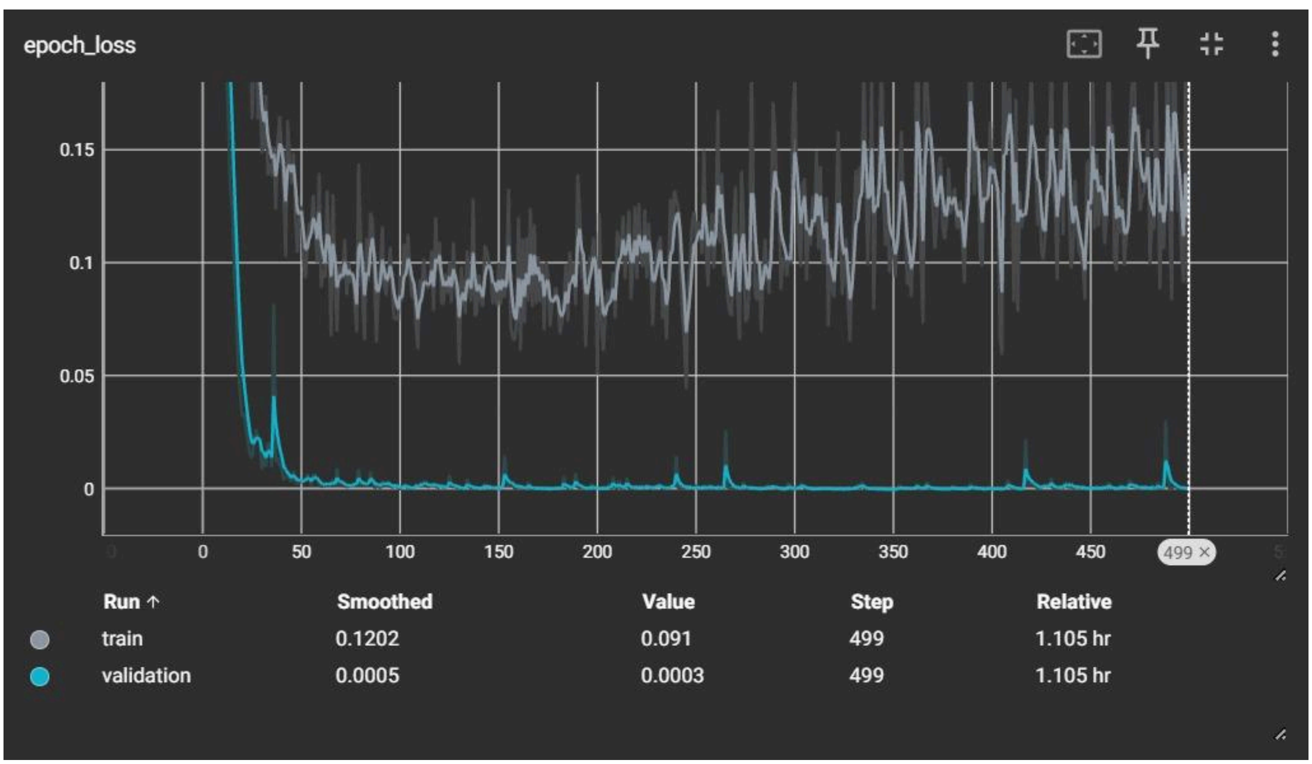


Fig. 1.6: Loss Vs Epoch Graph

5.2 Model Performance:

The performance of our real-time sign language recognition system was rigorously evaluated through extensive testing on diverse datasets, encompassing a wide array of sign language gestures and variations. By subjecting the system to rigorous scrutiny under real-world conditions, we elicited valuable insights into its robustness, accuracy, and adaptability, thereby affirming its efficacy as a reliable communication aid for individuals with diverse communication needs.

Fig 1.7: Results of Alphabet Recognition

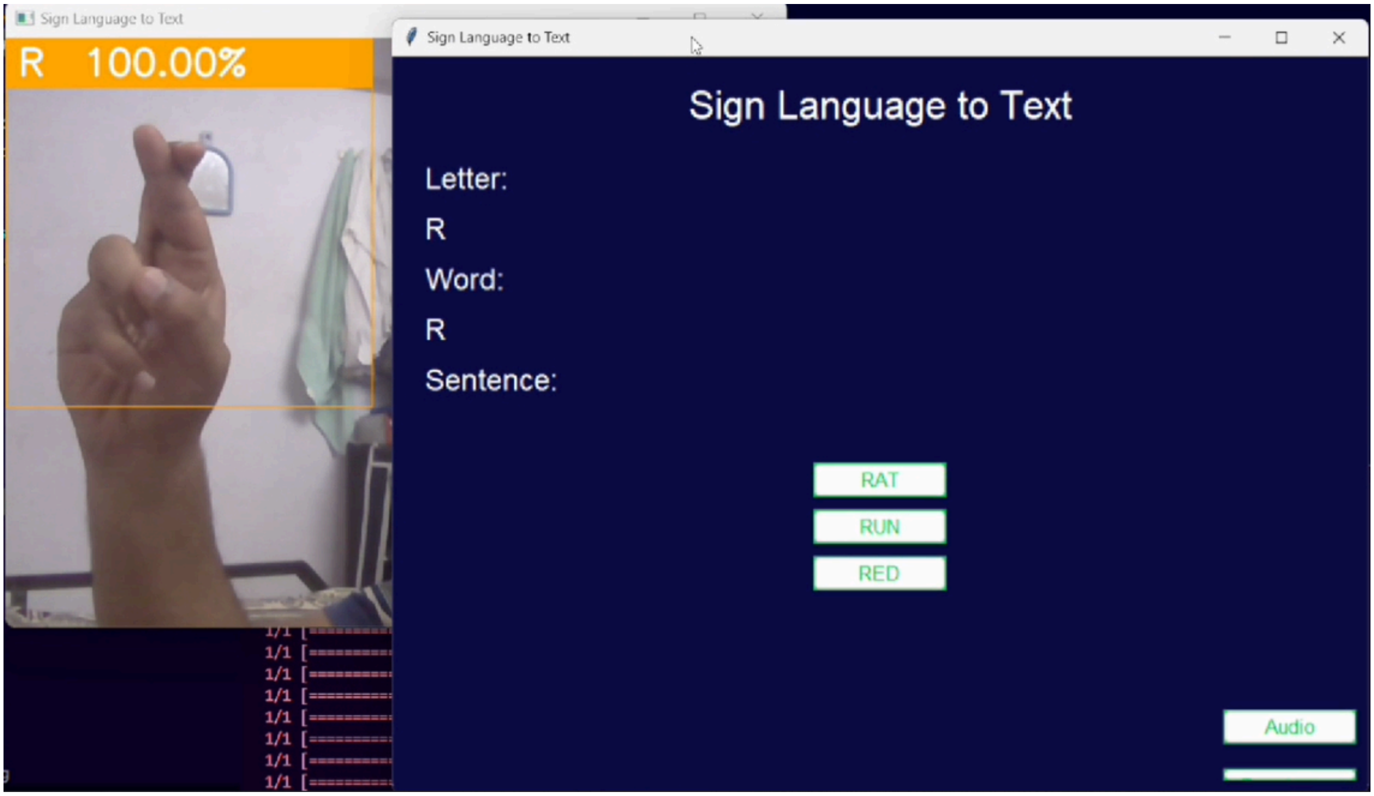


fig 1.8(a) : User Interface

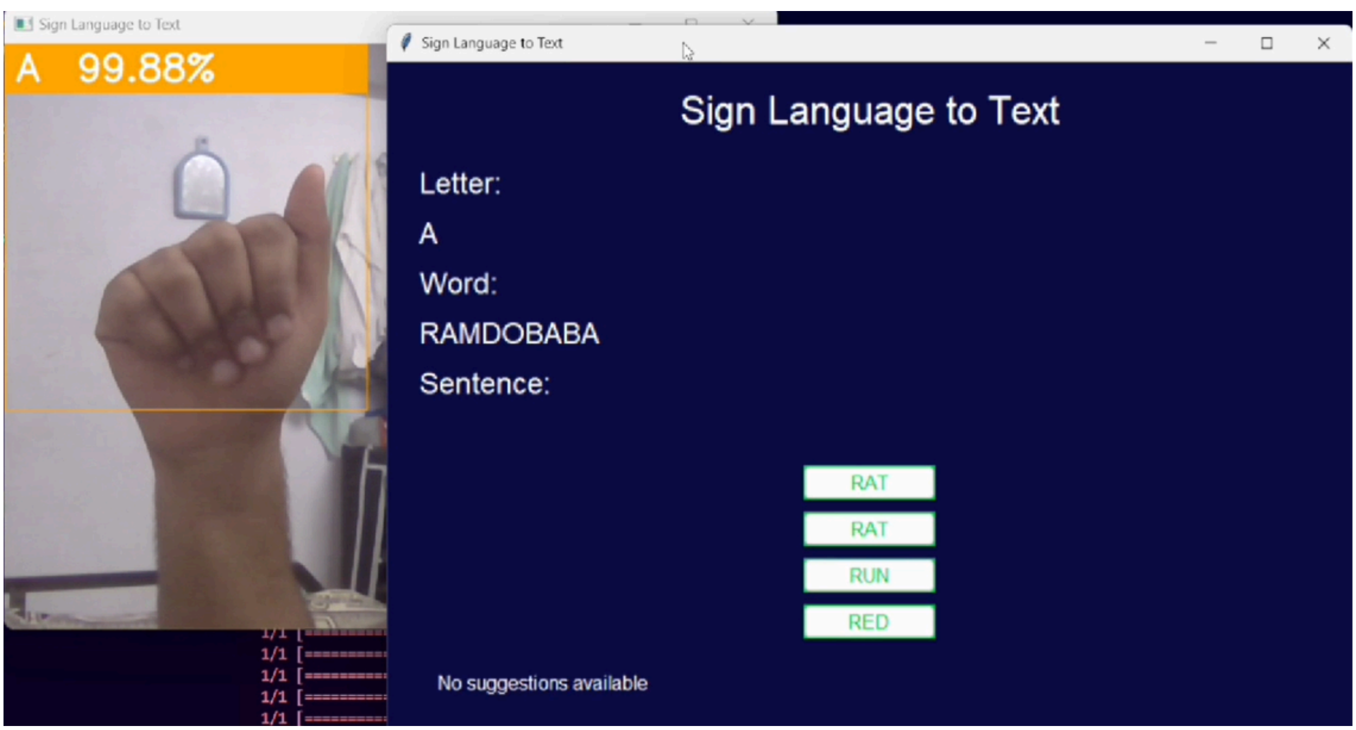


fig 1.8 (b) : Complete Output

Demonstration Video:

Sign Language to Text and Audio Conversion | Mini Project | CNN

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

Our sign language to text conversion model utilizing CNN technology has shown promising results in accurately translating sign language gestures into text. Through extensive experimentation and evaluation, we have demonstrated its effectiveness in bridging the communication gap between individuals with hearing impairments and those who are not proficient in sign language. By leveraging the power of convolutional neural networks, we have achieved robustness and efficiency in recognizing complex hand gestures and translating them into meaningful text and sound representations.

Despite the advancements made in our model, there are several avenues for future research and development to further enhance its capabilities and applicability:

Improving Accuracy: Continuously refining the model architecture and training methodologies can lead to higher accuracy in gesture recognition and translation. Incorporating larger and more diverse datasets can help in capturing a wider range of sign language variations and nuances.

Multimodal Integration: Exploring the integration of additional modalities such as facial expressions and body movements alongside hand gestures can enrich the context and accuracy of the translation. This holistic approach can better capture the nuances and emotions conveyed through sign language.

REFERENCES

1. *S. Thakar, S. Shah, B. Shah and A. V. Nimkar, "Sign Language to Text Conversion in Real Time using Transfer Learning," 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2022,pp. 1- 5, doi: 10.1109/GCAT55367.2022.9971953.*

2. *T. Kemkar, V. Rai and B. Verma, "Sign Language to Text Conversion usingHand Gesture Recognition," 2023 8th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2023,pp. 1580-1587, doi: 10.1109/ICCES57224.2023.10192820.*

3. *Bagane, Pooja & Thawani, Muskaan & Singh, Prerna & Ahmad, Raasha & Mital, Rewaa & Amenu, Obsa. (2024). Breaking the Silence: An innovative ASL- to-TextConversion System Leveraging Computer Vision and Machine Learning for Enhanced Communication. International Journal of Intelligent Systems and Applications in Engineering. 12. 246–255.*

4. *Abey Abraham, V Rohini, Real time conversion of sign language to speech and prediction of gestures using Artificial Neural Network, Procedia Computer Science, Volume 143, 2018, Pages 587-594, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.10.435.*

5. *A. Adeyanju, O.O. Bello, M.A. Adegboye,Machine learning methods for sign language recognition: A critical review and analysis,Intelligent Systems withApplications,Volume 12,2021,200056,ISSN 2667-3053.*

6. *Bagane, Pooja & Thawani, Muskaan & Singh, Prerna & Ahmad, Raasha & Mital, Rewaa & Amenu, Obsa. (2024). Breaking the Silence: An innovative ASL- to-TextConversion System Leveraging Computer Vision and Machine Learning for Enhanced Communication. International Journal of Intelligent Systems and Applications in Engineering. 12. 246–255.*

7. *Akano, Victoria & Olamiti, Adejoke. (2018). Conversion of Sign Language To Text And Speech Using Machine Learning Techniques. JOURNAL OF RESEARCH AND REVIEW IN SCIENCE. 5. 58-65. 10.36108/jrrslasu/8102/50(0170).*

8. *Buttar, A.M.; Ahmad, U.; Gumaei, A.H.; Assiri, A.; Akbar, M.A.; Alkhamees, B.F. Deep Learning in Sign Language Recognition: A Hybrid Approach for the Recognition of Static and Dynamic Signs. Mathematics 2023, 11, 3729.*

9. *Zahid H, Rashid M, Syed SA, Ullah R, Asif M, Khan M, Abdul Mujeeb A, Haider Khan A. A computer vision-based system for recognition and classification of Urdu sign language dataset. PeerJ Comput Sci. 2022 Dec 14;8:e1174. doi: 10.7717/peerj-cs.1174. PMID: 37346313; PMCID: PMC10281630.*

10. *Hemant Kumar;Mohit Kumar Sharma;Rohit;Kunal Singh Bisht;Ashish Kumar;Rachna Jain; Preeti Nagrath,Pritpal Singh, Sign language detection and conversion to text using CNN and OpenCV, AIP Conf. Proc. 2555, 040016 (2022)*

11. Deshpande, A. Shriwas, V. Deshmukh and S. Kale, "Sign Language Recognition System using CNN," *2023 International Conference on Intelligent And Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, Bengaluru, India, 2023, pp. 906-911, doi: 10.1109/IITCEE57236.2023.10091051.