Sign language recognition and translation

Submitted in partial fulfillment of the
requirements for the award of
Bachelor of Engineering degree in Computer Science and Engineering

By

KS PRAGALATHAN (Reg.No.39110785)
SAGI SANDEEP (Reg.No.39110865)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade "A" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHISALAI,
CHENNAI - 600119

NOVEMBER - 2022

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with ―A‖ grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600 119

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **KS PRAGALATHAN (39110785) and SAGI SANDEEP (Reg.No - 39110865)** who carried out the Project Phase-1 entitled **"AUDIO AND VIDEO TO SIGN LANGUAGE AND VICE-VERSA"** under my supervision from June 2022 to November 2022.

Internal Guide

Dr Ramya G Franklin, M.E., Ph.D.,

Head of the Department

Dr. L. LAKSHMANAN, M.E., Ph.D.,

Submitted for Viva voce Examination held on 20.04.2023

Internal Examiner **External Examiner**

DECLARATION

I, **KS PRAGALATHAN** (Reg.No-39110785) and **SAGI SANDEEP** (Reg.No-39110865), hereby declare that the Project Phase-1 Report entitled **"AUDIO AND VIDEO TO SIGN LANGUAGE AND VICE-VERSA"** done by me under the guidance of Dr Ramya G Franklin**, M.E, Ph.D.,** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in **Computer Science and Engineering**.

Sagi Sandeep

DATE: 27-10-2022

PLACE: Chennai **SIGNATURE OF THECANDIDATE**

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph.D., Dean**, School of Computing, **Dr. L. Lakshmanan M.E., Ph.D.,** Head of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide Dr Ramya G Franklin, **M.E, Ph.D.,** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my phase-1 project work.

I wish to express my thanks toall Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

Machines' ability to comprehend human actions and their meaning can be used in a wide range of situations. Sign language recognition is one area of research that has piqued curiosity. Deaf and dumb persons use sign language as a means of communication. They must rely on sign language translators to communicate with the rest of the world. The strategies employed in contemporary sign language recognition research are reviewed in this research. Data capture, preprocessing, segmentation, feature extraction, and classification are all stages of the methodologies that are examined. Each stage of the above-mentioned literature review is completed independently. Artificial neural networks, Support Vector Machines and Fuzzy Inference Systems are some of the most common categorization approaches used for recognition.

Sign Language Recognition is one of the most growing fields of research area. Many new techniques have been developed recently. Sign language is one of the well-structured code language, where each sign has some meaning allocated to it

As we know sign language is the only way for communication for differently abled people. The computer can be modified in such a way that it can decipher sign dialect to content arrange, and can help to remove

differences between the typical individuals and the differently abled people. These proposed frameworks which is able to recognize the different letter and sentence sets for Human-Computer interaction giving more exact outcomes about at slightest conceivable time. It'll not only be an

	TITLE	Page No.

advantage for differently abled people but can also be utilized by people belonging to both technical and non-technical background in different applications within the innovation field The main purpose of these papers is to convert the signs

performed by the dumb people into its corresponding audio and text. In order to do this the user has to perform his/her signs in front of the camera. The system would compare the recorded actions to the contents of its database to play the audio and show the text to understand what the user is actually trying to convey.

Improvement in Science and Technology aims to better world. The need for a comfortable and simplified life paves way to opportunities that can enhance the existing technology. The empowerment of differently abled will always stand as a priority. Disabled people face non-ending difficulties. Inability to speak is considered to be true disability. People with this disability use different modes to communicate with others, there are n number of methods available for their communication one such common method of communication is sign language. Sign language allows people to communicate with human body language

CHAPTER 1

INTRODUCTION

The hearing-impaired community uses sign language as a means of communication. The Rights of Persons with Disabilities (RPwD) Act 2016, which took effect on April 19, 2017, recognizes sign language as a form of communication that is particularly effective for communicating with people who are deaf or hard of hearing. It consists of a variety of hand formations, orientations, hand movements, and facial expressions that are used to transmit messages. Every gesture has a specific message. Sign language is not universal. Different sign languages, such as the American Sign Language (ASL), British Sign Language (BSL), Japanese Sign Language, and so on, can be found all over the world.

It is tough and expensive to find experienced and competent Interpreters on a regular basis. Non-deaf people, on the other hand, never try to learn sign language in order to communicate with deaf people. This leads to the isolation of deaf people. The distance between normal people and the deaf community can be bridged if a computer can be programmed to translate sign language into text format. It can be used to generate voice, or text, allowing the deaf to gain independence.

The organization of the paper is as follows: Section II gives information about the previous work done on sign languages. Section III focuses on the sign language approaches. Section IV describes the overview of the sign language recognition system. Section V contains discussion and conclusion

Sign Language is the most natural and expressive way for the dumb and deaf people. Normal people, never try to learn the sign language for communicating with the hearing impaired people. This leads to separation of the deaf people. It has been estimated that there are between 0.9 and 14 million hearing impaired in India and perhaps ″one of every five people who are deaf in the world, lives in India″, making it the country with the largest number of Deaf, and perhaps also the largest number of sign language users. But if the system can be programmed in such a way that it can convert sign language to text and speech format such that, the difference between the deaf people and the normal community can be reduced.

A sign language is a way of communicating by using the hands and other parts of the body. The main idea is to convert sign language to text/speech. The framework assist speech impaired to communicate with the society using sign language. This leads to the elimination of the mediator who commonly acts as a medium of 194 translation.This would contain a user friendly environment for the user by providing speech text output for a sign gesture input.

The advancement in technology provides us a platform to explore and design a real time sign language recognition system. This system recognizes 26 hand gestures from Indian Sign Language by hand gesture for tech generation. Webcam is used to capture the gestures. Using color model, the signs are processed for feature extraction. The extracted features are compared by using pattern matching algorithm. The features are compared with testing database to calculate the sign recognition. Eventually, recognized hand gesture is converted into text. This system provides an opportunity for deaf and dumb people to easily communicate with normal people without the need of an interpreter.

CHAPTER 2

LITERATURE SURVEY

Starner and Pentland provided one of the early studies on sign language recognition. They demonstrated a real-time hidden Markov model-based system that detected sentence level American Sign Language (ASL) movements with the help of a webcam. They described two experiments: the first uses a desk-mounted camera to view the user, while the second uses a camera embedded in the user's cap.

The authors in paper presented a system for the automatic translation of the gestures of the manual alphabets in Arabic Sign Language. This system made use of images of the gesture as input which were then processed and converted into a set of features that comprised of some length measures which indicated the fingertip's position. The subtractive clustering algorithm and the least-squares estimator were used for classification. The system achieved an accuracy of 95.55%.

In Nadia R. Albelwi and Yasser M. Alginahi proposed a real-time Arabic Sign Language system where a video camera was used to capture real-time video as an input to the system. The authors used a Haar-like algorithm to track the hand in the video frames and applied preprocessing techniques like skin detection and size normalization to extract the region of interest. To obtain the feature vectors, Fourier Transformation is applied to the resultant images which are transformed into the frequency domain. The classification is performed using the k-Nearest Neighbor (KNN) algorithm and the system achieves an accuracy of 90.55%.

In Balakrishnan, G., P. S. Rajam, et al., proposed a system that converts a set of 32 combinations of the binary number, which represents the UP and DOWN positions of the five fingers into decimal. The binary numbers are first converted into a decimal form by using the binary-decimal conversion algorithm and then the decimal numbers are converted to their corresponding Tamil letters. Static images of the gesture were used as the input to the system where a canny-edge detection algorithm was applied to extract the edges of the palm and Euclidean Distance was applied to identify the position of

The authors in paper proposed a system employing bspline approximation to develop a novel vision-based recognition system for Indian sign language alphabets and digits. By using the Maximum Curvature Points (MCPs) as Control points, their technique approximates the extracted boundary from the region of interest to a B-Spline curve. The B-spline curve is then smoothed iteratively, resulting in the extraction of Key Maximum Curvature Points (KMCPs), which are the major contributors to the gesture shape. As a result, the spatial coordinates of the KMCPs in the 8 Octant Regions of the 2D Space that are given for classification yield a translation and scale-invariant feature vector. The accuracy of numbers was 93.2 percent, and the accuracy of alphabets was 91.83 percent.

In paper the authors proposed a system that can recognize and convert ISL gestures from a video feed into English voice and text. They did this by segmenting the shapes in the video stream using several image processing techniques such as edge detection, wavelet transform, and picture fusion. Ellipsoidal Fourier descriptors were used to extract shape features, while PCA was utilized to optimize and reduce the feature set. The fuzzy inference system was used to train the system, which resulted in a 91% accuracy. In paper, the authors suggested a method for automatically recognizing Indian sign language gestures. The proposed method employs digital image processing techniques and uses the YCbCr color space for hand detection, with the input image being transformed beforehand. Distance transformation, Projection of distance transformation coefficients, Fourier descriptors, and feature vectors are some of the techniques used to extract the features. An artificial neural network was used to classify the data, and the recognition rate was 91.11 percent.

2.1 SIGN RECOGNITION SYSTEM APPROACHES:

Sign language recognition system is classified into two categories: Sensor-based approach and Vision-based approach. Sensor-based gloves are employed in the first approach. Different sensors are used to collect data from the gesture. After then, the data is examined, and conclusions are drawn based on the findings. The problem of this approach is that the sensor glove must be carried with you at all times, i.e., the individual must wear the sensor as well as the glove for the system to work, and sensor-based gloves are fairly expensive. Cameras are utilized to capture images of the gesture as an input for the recognition system in the vision based approach. To detect the gesture, this system employs a variety of image processing algorithms. The vision-based sign recognition systems are further categorized into appearance-based and 3D model-based

Images or videos are utilized as inputs in appearance-based recognition, and features are derived from them. These characteristics are then matched to photos that have already been saved. The hand's 3D descriptors are used in the 3D model-based method. This method entails looking for the kinematic parameters that correspond to a 2D projection of a 3D model of a hand and an edge-based image.

Table -1: Comparison of Various Sign Language Recognition Systems

Ref No.	Input	Description	Feature Vector	Classification	Recognition Rate
[3]	Images	Processed and converted the images into a set of features that comprised of some length measures which indicated the fingertip's position.	A priori model of skin color	subtractive clustering algorithm and the least-squares estimator	95.55%
[4]	Videos	Haar-like algorithm to track the hand in the video frames and preprocessing techniques like skin detection and size normalization to extract the region of interest	Fourier Transformation	k-Nearest Neighbor (KNN)	90.55%
[6]	Images	Used the Maximum Curvature Points (MCPs) as Control points, to approximate the extracted boundary from the region of interest to a B-Spline curve.	A novel method based on B-spline approximation	Support Vector Machine (SVM)	Numbers-93.2% Alphabets-91.83%
[7]	Video	Segmented the shapes in the video stream using several image processing techniques such as edge detection, wavelet transform, and picture fusion.	Elliptical Fourier Descriptors	Fuzzy Inference System	91%
[8]	Images	Employs digital image processing techniques and uses the YCbCr color space for hand detection,	Distance transformation, Projection of distance transformation coefficients, Fourier descriptors, and feature vectors	Artificial Neural Network	91.11%

Figure 2.1

CHAPTER 3

REQUIREMENT ANALYSIS

3.1 SOFTWARE REQUIREMENTS SPECIFICATION DOCUMENT

Software Implemented through Python

A desktop application is implemented using python programming language.

Python this includes libraries such as pyaudio to convert speech to text.

- Python 2.7.x or above is preferred.
- Pycharm community edition compiler.
- Operating System - Ubuntu (Linux) or Windows.
- ISL/ASL data sets from google.

3.2 PROJECT MANAGEMENT PLAN

We are going to use machine learning sign language applications to manage our project and to ensure that work is completed on time.

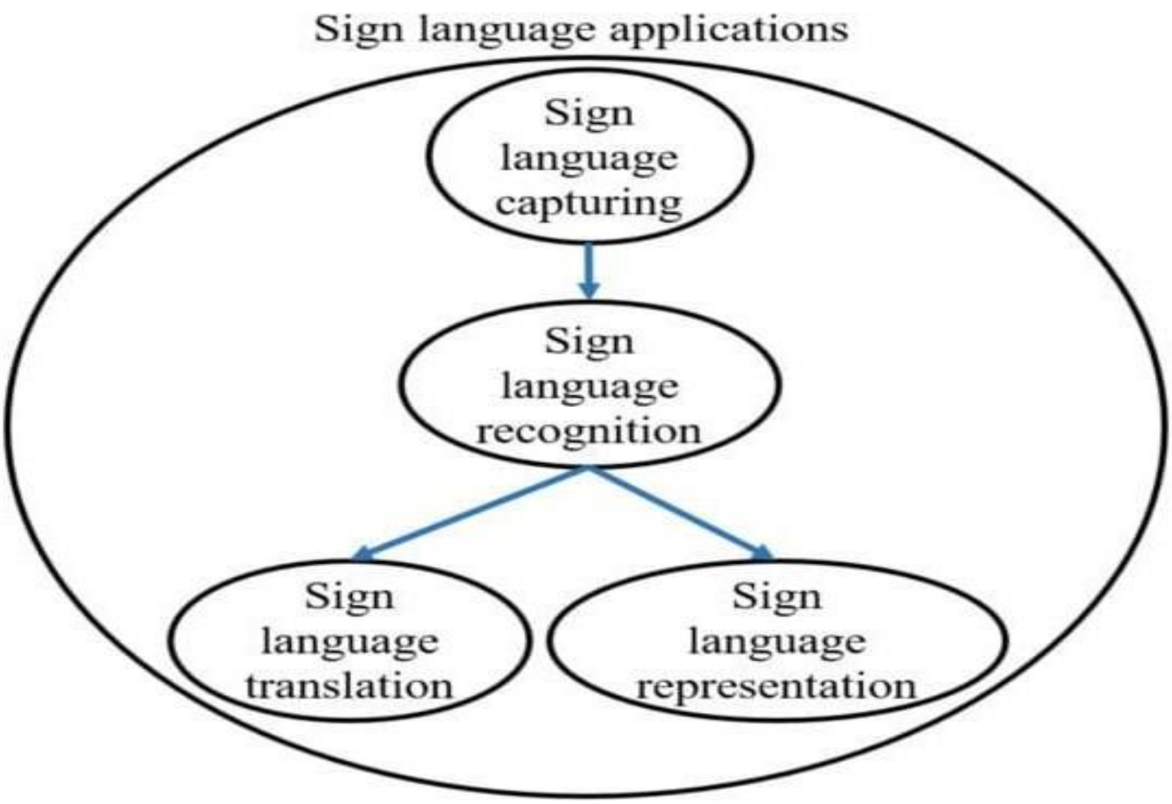

Figure 3.1

STEP 1: SIGN LANGUAGE CAPTURING

Sign language capturing involves the recording of sign gestures using appropriate sensor setups. The purpose is to capture discriminative information from the signs that will allow the study, recognition and 3D representation of signs at later stages.

STEP 2: SIGN LANGUAGE RECOGNITION

Sign language recognition (SLR) is the task of recognizing sign language glosses from video streams. It is a very important research area since it can bridge the communication gap between hearing and Deaf people, facilitating the social inclusion of hearing-impaired people.

STEP 3: SIGN LANGUAGE TRANSLATION

Sign Language Translation is the task of translating videos with sign language into spoken language by modeling not only the glosses but also the language structure and grammar. It is an important research area that facilitates the communication between the Deaf and other communities.

STEP 4: SING LANGUAGE REPRESENTATION

The automatic and realistic sign language representation is vital for each sign language system. The representation of a sentence in sign language instead of a plain text can make the system friendlier and more accessible to the members of the deaf community.

CHAPTER 4

OVERVIEW OF THE INDIAN SIGN LANGUAGE RECOGNITION SYSTEM

The goal of a sign language recognition system is to accurately recreate speech or text based on a given sign. Table 1 shows the comparison of various Sign Language recognition systems while Table 2 describes the positives and negatives of the various sign language systems. The important steps involved in the sign language recognition system include preprocessing, feature extraction, and classification.

4.1Pre-processing

To extract the region of interest from the training images, a pre-processing step is performed (ROI). If only hand motions are taken into account, the ROI can be hands, or face and hands if facial gestures are also taken into account. Filtering, picture enhancement, image resizing, segmentation, and morphological filtering are common pre-processing steps. Any of the generally used methods for filtering and image enhancement can be employed. Otsu's thresholding, background subtraction, skin color-based segmentation, and motion-based segmentation are some of the algorithms utilized for segmentation. The test photos or videos are also pre-processed to extract the region of interest during the testing phase.

4.2Feature Extraction

The feature vectors obtained from this step are the inputs to the classifier, so feature extraction is one of the most important steps in sign language recognition.

The feature extraction approaches should be able to discover shapes accurately regardless of changes in lighting levels, orientation, and size of the object in a video/image. Wavelet decomposition, Haar wavelets, Haar-like features, orientation histogram, scale invariant feature transform, Fourier descriptors, and other methods can be used to produce the features. The classifier is then trained using the feature vector acquired using any of the feature extraction methods.

4.3Classification

In order to classify the input signals into distinct classes, a classifier is required in sign language recognition. During the training phase, the feature vector obtained from the training database is used to train the classifier. When a test input is presented, the trained classifier recognizes the sign's class and displays or plays the appropriate text or sound. Images or videos can be used as test inputs.

Machine learning techniques for classification are classified into two types: supervised and unsupervised. Supervised machine learning is a method of teaching a computer to detect patterns in input data that can subsequently be used to predict future data. Supervised machine learning applies a collection of known training data to labeled training data to infer a function. Unsupervised machine learning is used to make inferences from datasets with no tagged response. Because no labeled response is offered to the classifier, there is no reward or punishment weightage to which classes the data is meant to go. Hidden Markov Models (HMM), Artificial Neural Networks (ANN), Multiclass Support Vector Machines (SVM), Fuzzy Systems, and K Nearest Neighbor (KNN) are some of the most often used classifiers. The recognition rate is used to evaluate the classifier's performance.

Convolutional Neural Network

A multilayered feedforward neural network is referred to as a convolutional neural network(CNN). It's one of the most often used approaches for categorizing photographs.

There are numerous convolution layers, a pooling layer, and an activation function in it.

The fully connected layer is at the very end. The input to the CNN is a vector representation

of the image, and the predicted label is the output. The completely linked layer has the

same number of neurons as classes. It also has a loss function, which determines the algorithm's effectiveness

K-Nearest Neighbor

The k nearest neighbor algorithm is a supervised learning system that uses feature space to classify objects. The nearest-neighbor method is used bythe KNN method to classify each row of data in the sample into one of the training groups. A set of features obtained during the training phase utilizing a variety of training photos makes up the category. The fundamental goal of classification is to discover the best matching features vector for the new vector among the reference features.

4.4 ARCHITECTURE / OVERALL DESIGN OF PROPOSED SYSTEM

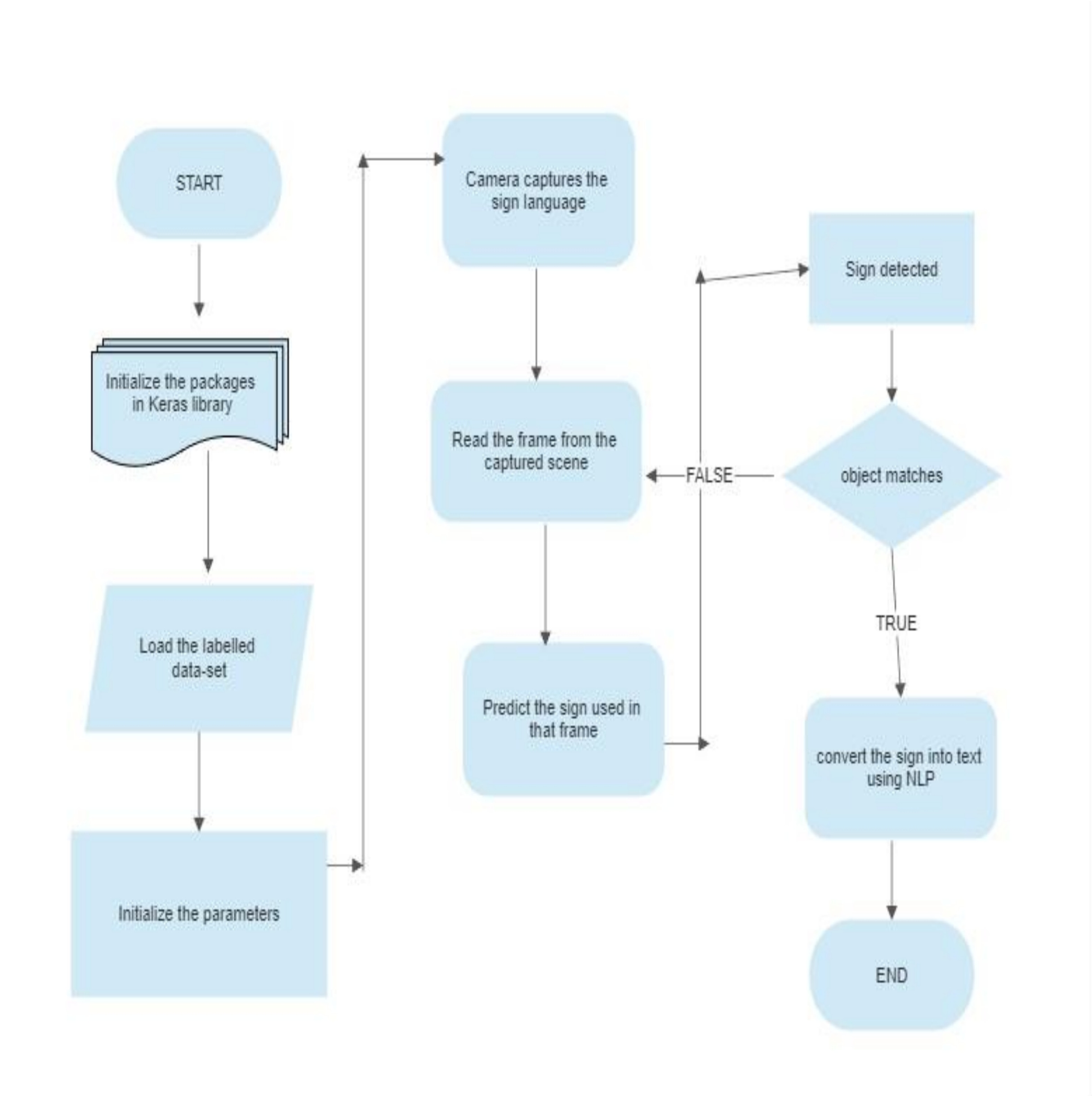

Figure 4.1

FLOWCHART

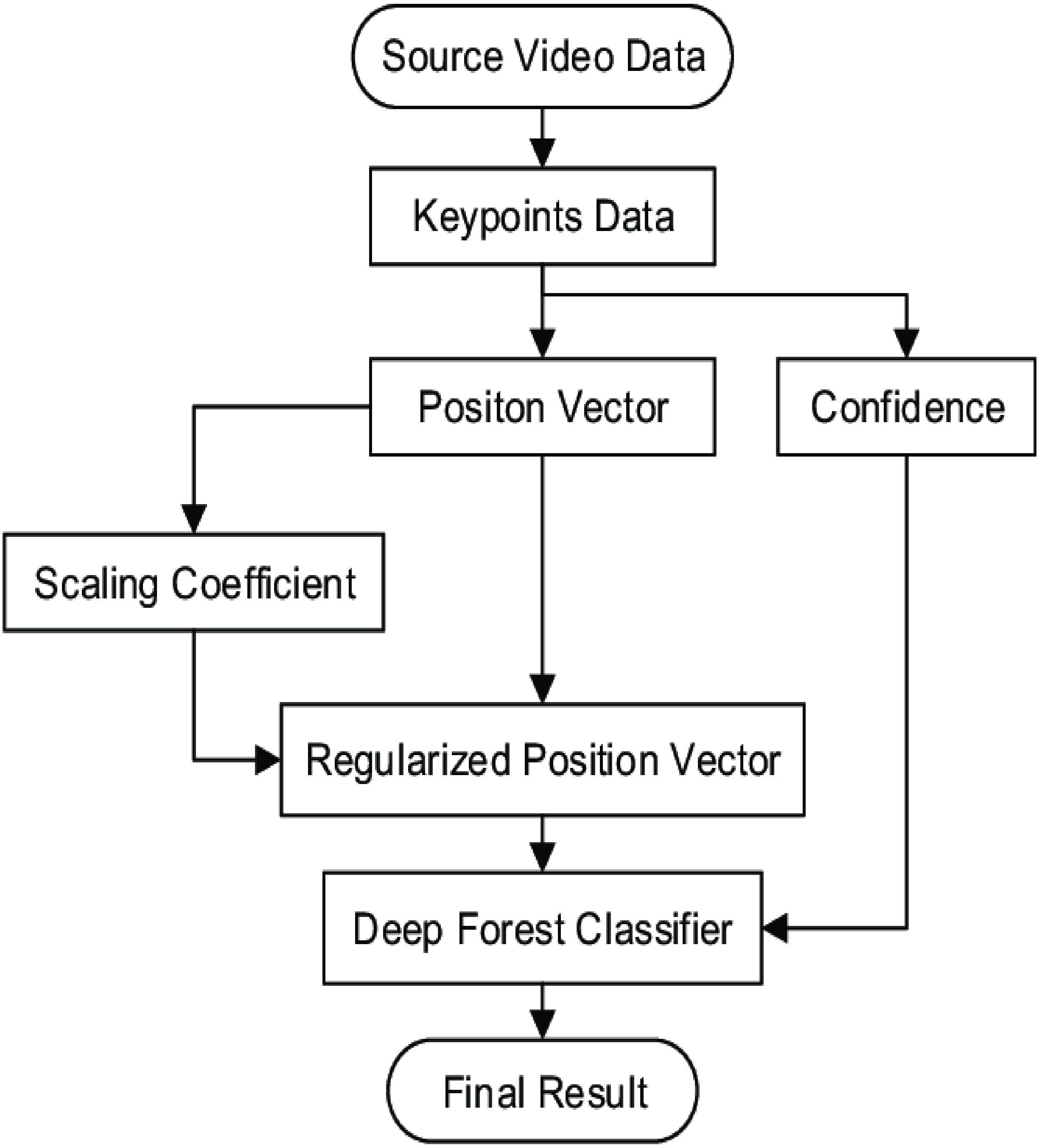

Figure 4.2

The below image is designed to provide objective and subjective comparisons of AI technologies and DNN architectures for sign language as seen from the perspective and the experience of the authors in the field. More specifically, [a] presents and compares the characteristics of the different AI technologies for sign language. Volume of works is used to measure the number of published papers for each sign language technology and it is calculated based on the results of the query search in the databases. Finally, future potential is used to express the view of the authors on which sign language technology has the most potential to deliver future research works.

As Below image shows that Radar charts showcasing the findings of this survey regarding

(a) the literature methods for CSLR, ISLR and SLP and

(b) the characteristics of each AI sign language technology.

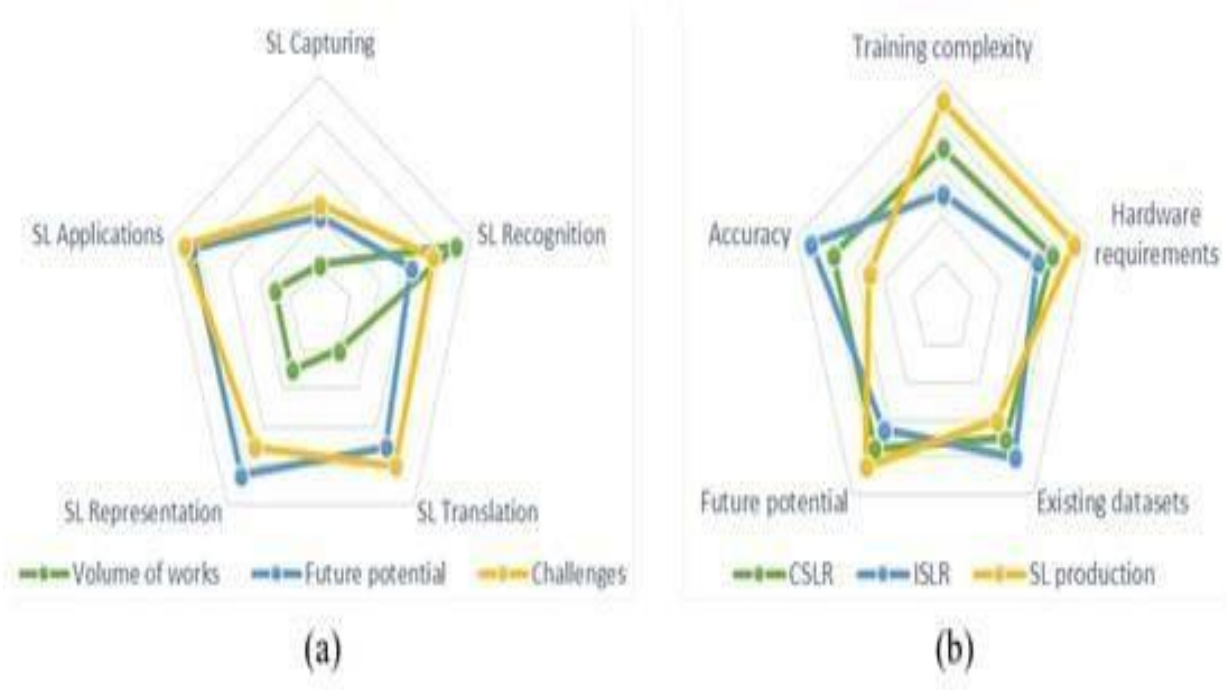

Figure 4.3

[a] it can be seen that most existing works deal with sign language recognition, while sign language capturing and translation methods are still not thoroughly explored. It is strongly believed that these research areas should be explored more in future works. Furthermore, it is assumed that there is still great room for improvement for applications, especially mobile ones, that can assist the Deaf community.

Regarding future directions, improvements can still be achieved in the accuracy of sign language recognition and production systems. In addition, advances should be made in the extraction of robust skeletal features, especially in the presence of occlusions, as well as in the realism of avatars. Finally, it is crucial to develop fast and robust sign language applications that can be integrated in the everyday life of hearing-impaired people and facilitate their communication with other people and services.

[b] draws a comparison between various DNN architectures in terms of the performance of the proposed networks, hardware requirements for inference and training of the proposed networks, scope for improvement based on the performance gains and the volume of works, computational complexity during training and the number of recorded datasets that are currently available. Except for the existing datasets, whose values are based on a search for publicly available datasets, all other metrics presented in the chart of [b] are calculated based on the study of the review papers and the opinions and experience of the authors. As it can be observed, ISLR methods have high accuracy with small hardware requirements but such methods have been extensively explored resulting in limited future potential. On the other hand, CSLR and SLP methods have high hardware and training requirements, as well as demonstrate significant future potential as there is still great room for improvements in future research works.

5. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, the broad spectrum of AI technologies in the field of sign language is covered. Starting from sign language capturing methods for the collection of sign language data and moving on to sign language recognition and representation techniques for the identification and translation of sign language, this review highlights all important technologies for the construction of a complete AI-based sign language system. Additionally, it explores the in-between relations among the AI technologies and presents their advantages and challenges. Finally, it presents ground breaking sign language applications that facilitate the communication between hearing-impaired and speaking people, as well as enable the social inclusion of hearing-impaired people in their everyday life. The aim of this review is to familiarize researchers with sign language technologies and assist them towards developing better approaches.

In the field of sign language capturing, it is essential to select an optimal sensor for capturing signs for a task that highly depends on various constraints (e.g., cost, speed, accuracy, etc.). For instance, wearable sensors (i.e., gloves) are expensive and capture only hand joints and arm movements, while in recognition applications, the user is required to use gloves. On the other hand, camera sensors, such as web or smartphone cameras, are inexpensive and capture the most substantial information, like the face and the body posture, which are crucial for sign language.

Concerning CSLR approaches, most of the existing works adopt 2D CNNs with temporal convolutional networks or recurrent neural networks that use video as input. In general, 2D methods have lower training complexity compared to 3D architectures and produce better CSLR performance. Moreover, it is experimentally shown that multi-modal architectures that utilize optical flow or human pose information, achieve slightly higher recognition rates than unimodal methods. In addition, CSLR performance on datasets with large vocabularies of more than 1000 words, such as Phoenix-2014, or datasets with unseen words on the test sets, such as CSL Split 2 and GSL SD, is far from perfect. Furthermore, ISLR methods have been extensively explored and have achieved high recognition rates on large-scale datasets. However, they are not suitable for real-life applications since they are trained to detect and classify isolated signs on pre-segmented videos.

Sign language translation methods have shown promising results although they are not exhaustively explored. The majority of the SLT methods adopt architectures from the field of neural machine translation and video captioning. These approaches are of great importance, since they translate sign language into spoken counterparts and can be used to facilitate the communication between the Deaf community and other groups. To this end, this research field requires additional attention from the research community.

Sign language representation approaches adopt either 3D avatars or video generation architectures. 3D animations require manual design of the movement and the position of each joint of the avatar, which is very time-consuming. In addition, it is extremely difficult to generate smooth and realistic animations of the fine grained movements that compose a sign, without the use of sophisticated motion capturing systems/technologies that employ multiple cameras and specialised wearable sensors.

On the other hand, recent deep learning methods for sign language production have shown promising results at synthesizing sign language videos automatically. Besides, they can generate realistic videos using a reference image or video from a human, which are also preferable from the Deaf community instead of avatars.

Regarding the sign language applications, they are mostly developed to be integrated in a smartphone operating system and perform SL translation or recognition. A discrete category is the educational oriented applications, which are very useful for anyone with little or no knowledge of sign language. In order to create better and more easily accessible applications, the research should focus on the development of more robust and less computational expensive AI models, along with the further improvement of the existing software for integration of the AI models into smart devices.

As per the literature review of all the above papers, they have used different techniques like sensor-based technology, image processing techniques, vision-based technology, discrete wavelet transformation technique, and machine learning using python for conversion of sign language into text and speech.These system consists of web camera which is used to capture the gestures. The signs are processed for feature extraction. The extracted features are compared by using pattern matching algorithm. The features are compared with testing database to calculate the sign recognition. Eventually, recognized hand gesture is converted into text and speech. This system provides an opportunity for deaf and dumb people to easily communicate with normal people without the need of an interpreter.

6. SCREENSHOTS

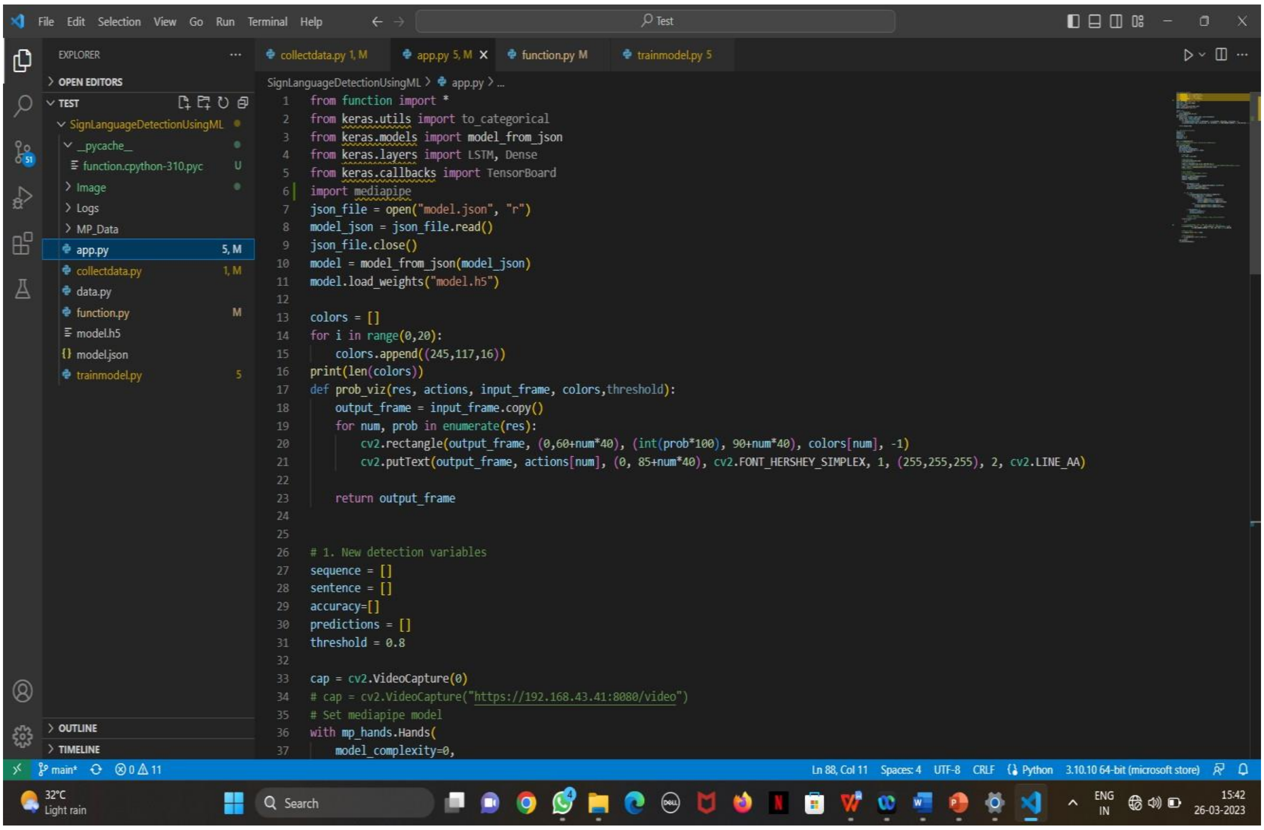

Figure 6.1

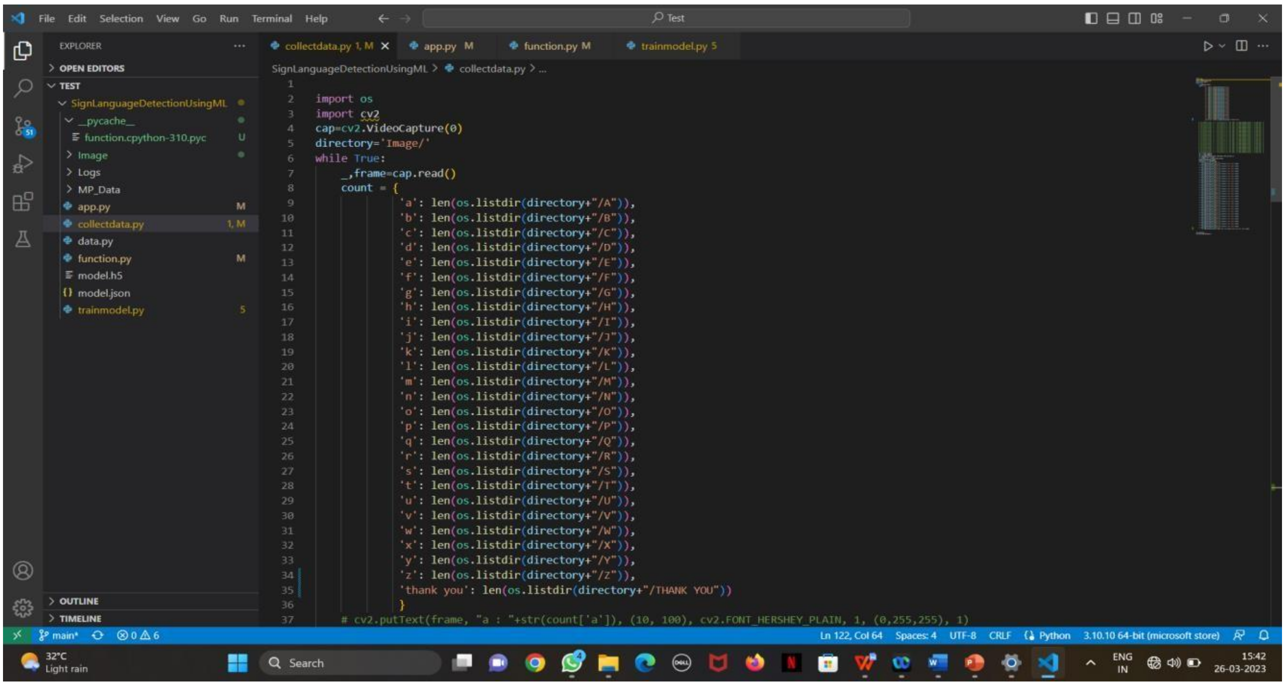

Figure 6.2

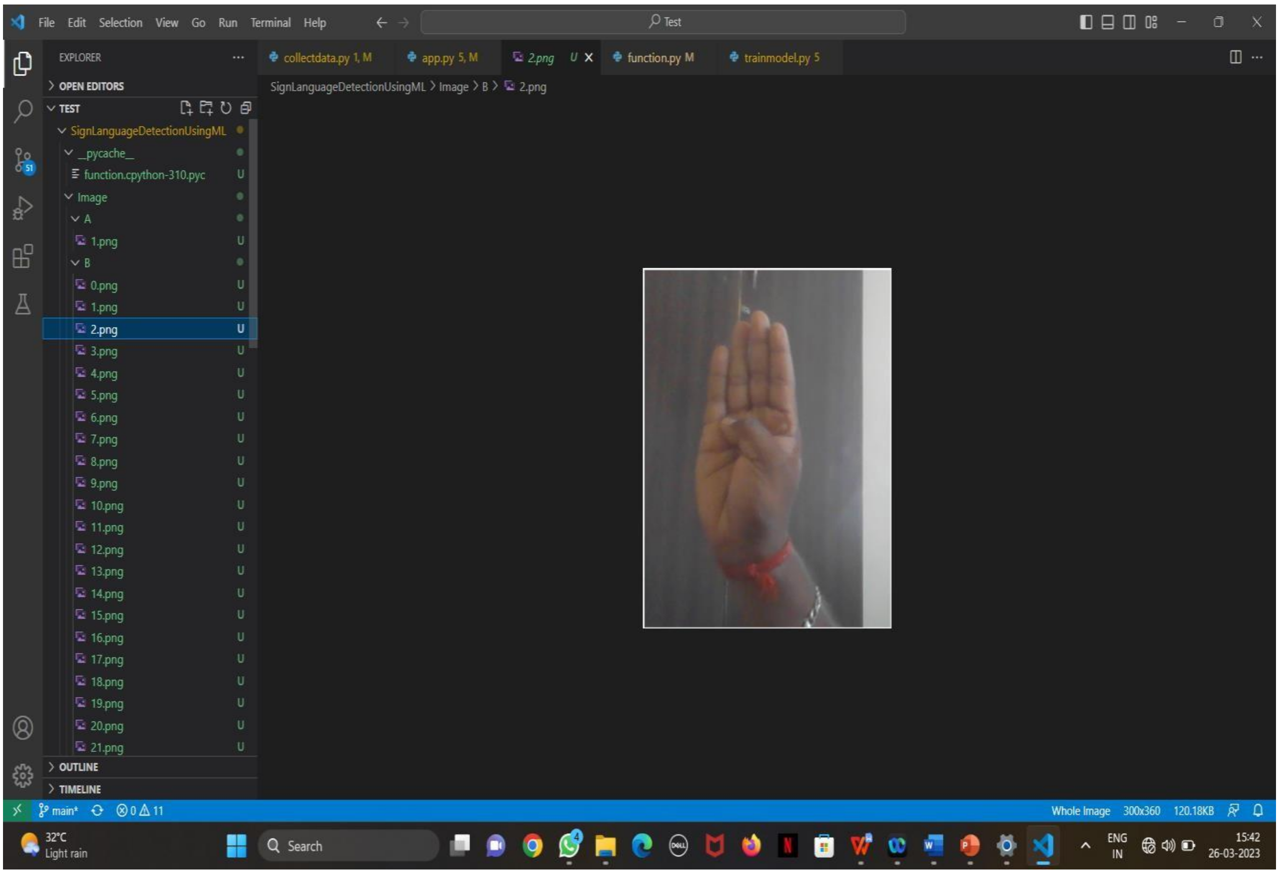

Figure 6.3

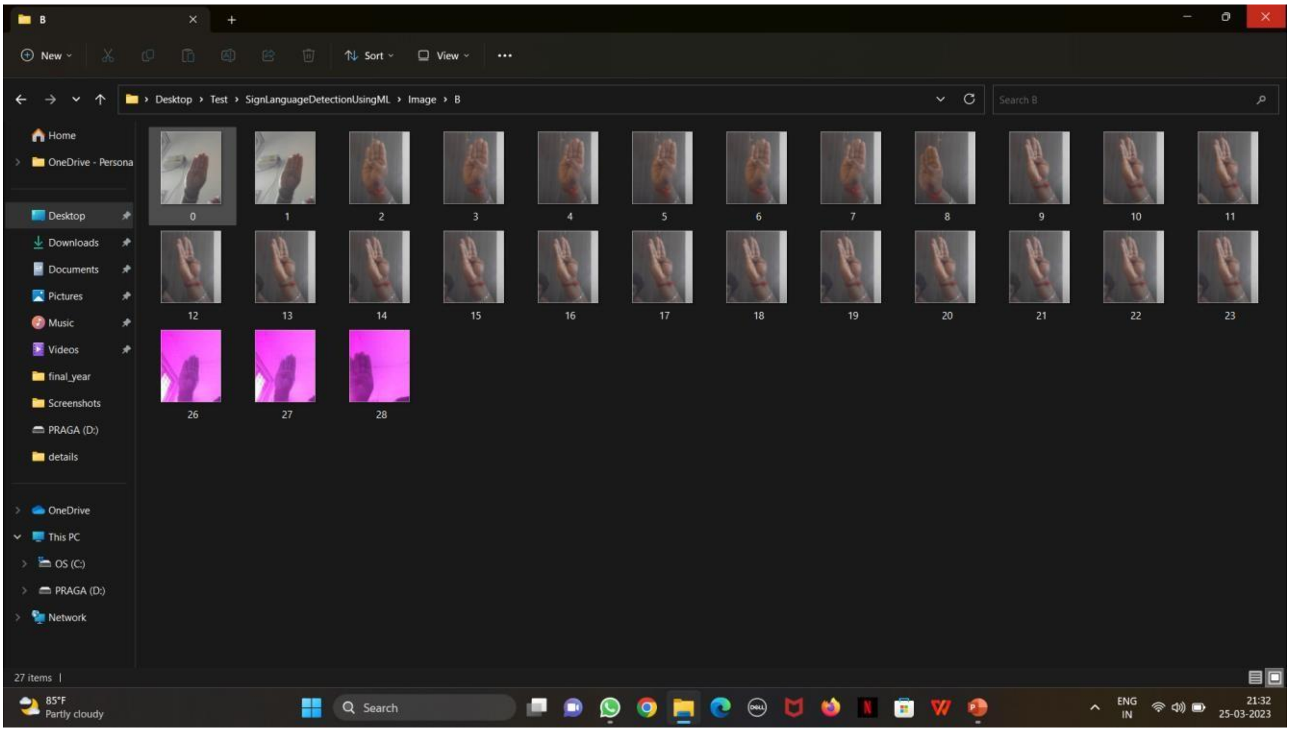

Figure 6.4

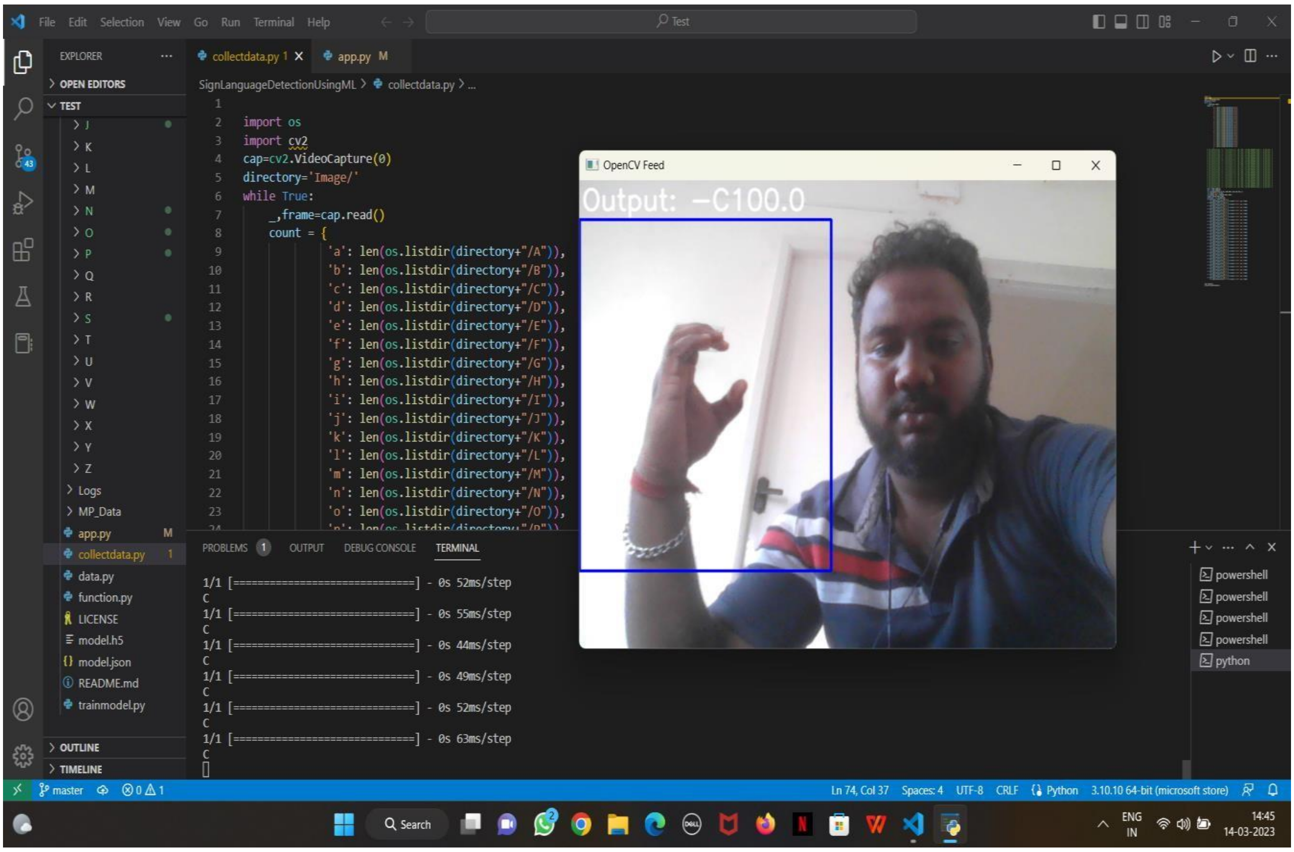

Figure 6.5
(Output)

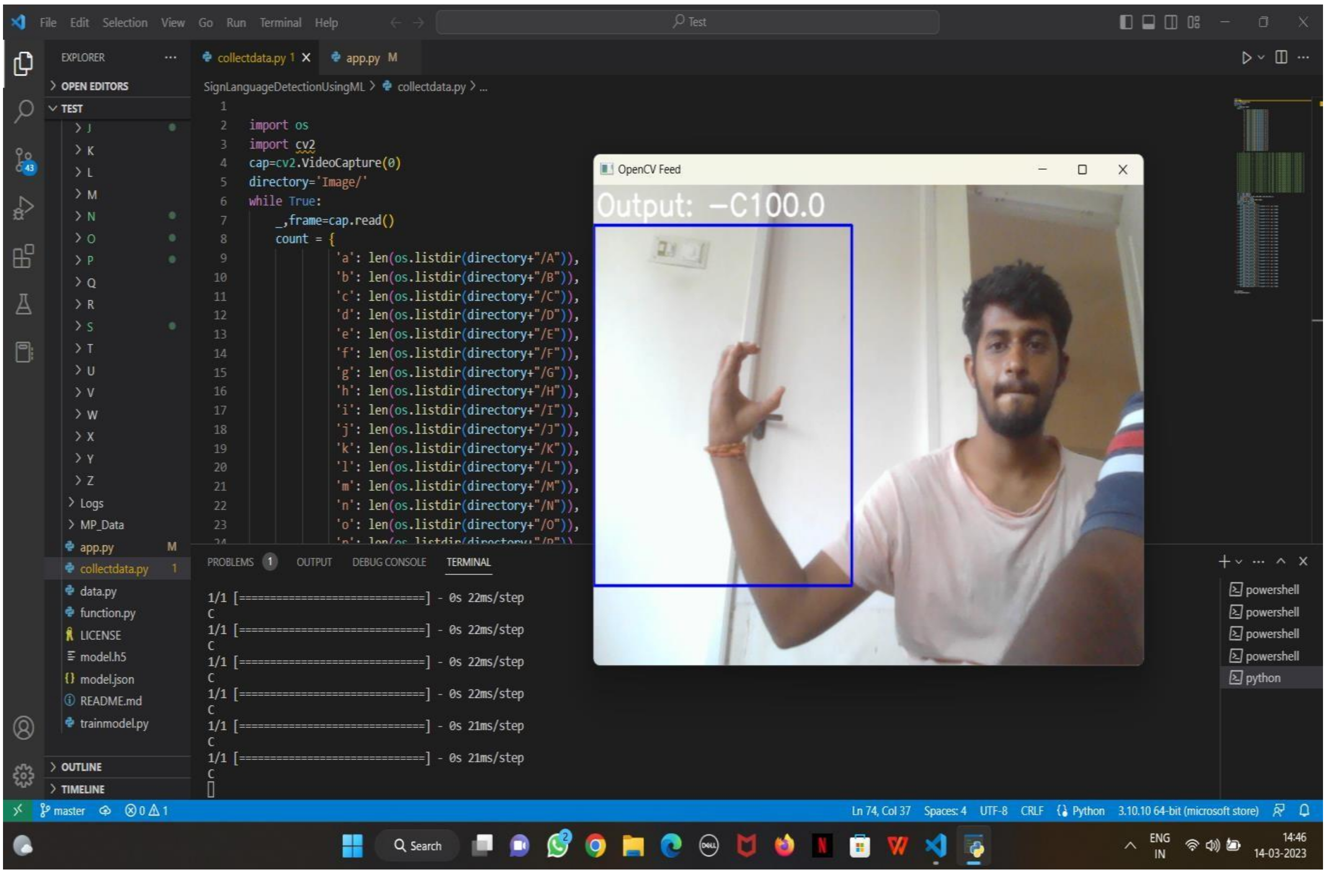

Figure 6.6

REFERENCES

- Noor Adnan Ibraheem, Rafiqul Zaman Khan, 2018, Survey on Various Gesture Recognition Technologies and Techniques, International Journal of Computer Applications, Volume 50, No. 7
- G. R. S. Murthy & R. S. Jadon, 2020. A Review of Vision Based Hand Gestures Recognition, International Journal of Information Technology and Knowledge Management.
- Al-Jarrah, A. Halawani, "Recognition of gestures in Arabic sign language using neuro-fuzzy systems," The Journal of Artificial Intelligence (2020)

- Nadia R. Albelwi, Yasser M. Alginahi, “Real-Time Arabic Sign Language (ArSL) Recognition International Conference on Communications and Information Technology 2018
- Rajam, Subha & Ganesan, Balakrishnan. (2019). Real time Indian Sign Language Recognition System to aid deaf-dumb people. International Conference on Communication Technology Proceedings, ICCT. 10.1109/ICCT.2011.6157974.
- Geetha M., Manjusha U. C. 2018 A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation, International Journal of Computer Science and Engineering (IJCSE)