A

SYNOPSIS

ON

"SIGN LANGUAGE TO TEXT AND SPEECH CONVERSION USING CNN"

For the award of degree of

Bachelor Of Technology

In

Computer Science and Engineering

By

ABHAY DHIMAN

SCHOOL OF COMPUTER SCIENCE & ENGINEERING
DEPARTMENT OF B.TECH. (CSE)
GOVT. P.G. COLLEGE DHARAMSHALA
HIMACHAL PRADESH

2020-2024

University Roll No: -20010603001 **Class Roll No.: -20539**

TABLE OF CONTENTS

S.NO.	CONTENT	PAGE NO.
1	Introduction	2-4
2	Objective of Project	4-6
3	Project Description	6-8
4	Project Category	8-9
5	Tools and Environments used	9
6	Project Planning	10-12
7	System Diagrams	13-16
8	Future Scope of the Project	17-19
9	References	20

1. INTRODUCTION

Sign language is one of the oldest and most natural form of language for communication, hence we have come up with a real time method using neural networks for finger spelling based American sign language. Automatic human gesture recognition from camera images is an interesting topic for developing vision. We propose a convolution neural network (CNN) method to recognize hand gestures of human actions from an image captured by camera. The purpose is to recognize hand gestures of human task activities from a camera image. The position of hand and orientation are applied to obtain the training and testing data for the CNN. The hand is first passed through a filter and after the filter is applied where the hand is passed through a classifier which predicts the class of the hand gestures. Then the calibrated images are used to train CNN.

American sign language is a predominant sign language Since the only disability D&M people have been communication related and they cannot use spoken languages hence the only way for them to communicate is through sign language. Communication is the process of exchange of thoughts and messages in various ways such as speech, signals, behaviour and visuals. Deaf and dumb(D&M) people make use of their hands to express different gestures to express their ideas with other people. Gestures are the nonverbally exchanged messages and these gestures are understood with vision. This nonverbal communication of deaf and dumb people is called sign language.

In our project we basically focus on producing a model which can recognise Fingerspelling based hand gestures in order to form a complete word by combining each gesture.

1.2 GESTURE RECOGNITION

The user's sign language gestures are recorded by cameras or other sensors. In order to recognize particular hand gestures, body movements, and facial expressions used in sign language, computer vision algorithms analyse these gestures.

1.3 GESTURE INTERPRETATION

The intended meaning of the signed message is deciphered using machine learning models or pattern recognition algorithms after the gestures have been identified.

1.4 TEXT GENERATION

The technology translates the sign language interpretation into text that may be shown on a screen or utilized in a variety of apps for additional processing.

1.5 TEXT TO SPEECH CONVERSION

The technique known as text-to-speech (TTS) transforms written text into spoken words. The voice synthesis method used by this technology involves the system analysing the written text and producing matching speech sounds. This is how it usually goes:

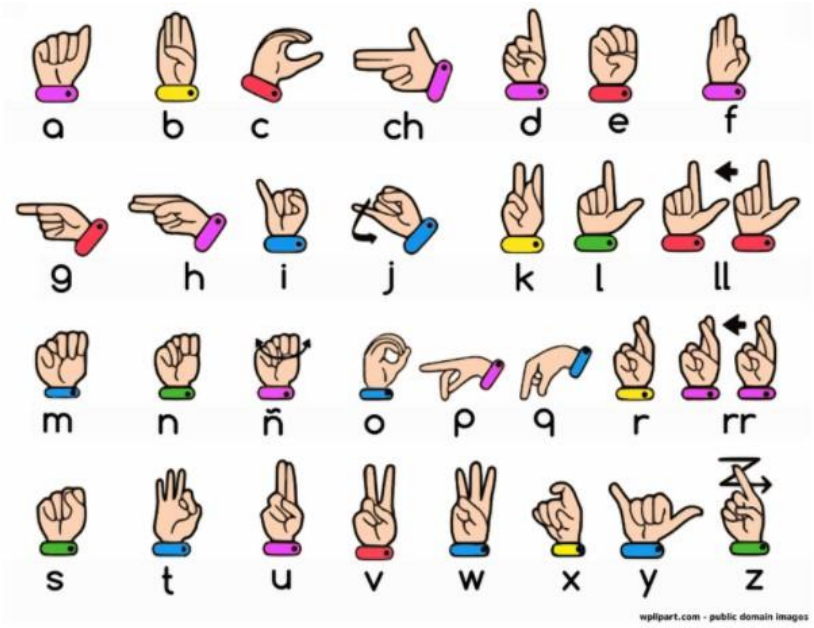
1.6 TEXT ANALYSIS

The words, punctuation, and sentence structure of the incoming text are identified through analysis. To produce more expressive and genuine voice, modern TTS systems can also tackle natural language processing tasks.

1.7 PHONEME GENERATION

The text is divided into phonemes, which are a language's smallest units of sound. Speech Synthesis:

The system assembles the phonemes into spoken words, phrases, and paragraphs using pre-recorded or created speech units, creating an audio representation of the original text. Audio Output: By playing the synthetic speech through speakers, headphones, or any other audio output device, users may hear the actual text spoken aloud. A powerful communication tool that allows deaf or hard of hearing persons to interact with hearing people who might not understand sign language is made possible by the combination of text-to-speech and sign language-to-text technologies. Applications for this technology included video relay services, communication tools, and live captioning. This Effective communication is essential in all aspects of life, and it is especially important for individuals who are deaf or hard of hearing. With the rising number of people suffering from hearing loss, it is crucial to find ways to bridge the communication gap between the hearing and non-hearing population. To address this issue, we present a new system for converting Sign Language into text format using computer vision and machine learning techniques. This system aims to provide an efficient and accessible solution for deaf and hard of hearing individuals to communicate with the hearing population. In the today's world, Communication is always having a great impact in every domain and how it is considered the meaning of thoughts and expressions that attract the researchers to bridge this gap for normal and deaf people. According to World Health Organization, by 2050 nearly 2.5 billion people are projected to have some degree of hearing loss and at least 700 million will require hearing rehabilitation. Over 1 billion young adults are at the risk of permanent, avoidable hearing loss due to unsafe listening practices. Sign languages vary among regions and countries, with Indian, Chinese, American, and Arabic being some of the major sign languages in use today. This system focuses on Indian Sign Language and utilizes the Media Pipe Holistic Key points for hand gesture recognition.



2. OBJECTIVE OF PROJECT

More than 70 million deaf people around the world use sign languages to communicate. Sign language allows them to learn, work, access services, and be included in the communities.
It is hard to make everybody learn the use of sign language with the goal of ensuring that people with disabilities can enjoy their rights on an equal basis with others.
So, the aim is to develop a user-friendly human computer interface (HCI) where the computer understands the American sign language This Project will help the dumb and deaf people by making their life easy.

2.1 Objective

To create a computer software and train a model using CNN which takes an image of hand gesture of American Sign Language and shows the output of the particular sign language in text format converts it into audio format.

2.2 Scope

This System will be Beneficial for Both Dumb/Deaf People and the People Who do not understands the Sign Language. They just need to do that with sign Language gestures and this system will identify what he/she is trying to say after identification it gives the output in the form of Text as well as Speech format.

2.3 Data Acquisition

The different approaches to acquire data about the hand gesture can be done in the following ways:

It uses electromechanical devices to provide exact hand configuration, and position. Different glove-based approaches can be used to extract information. But it is expensive and not user friendly.

In vision-based methods, the computer webcam is the input device for observing the information of hands and/or fingers. The Vision Based methods require only a camera, thus realizing a natural interaction between humans and computers without the use of any extra devices, thereby reducing costs. The main challenge of vision-based hand detection ranges from coping with the large variability of the human hand's appearance due to a huge number of hand movements, to different skin-color possibilities as well as to the variations in viewpoints, scales, and speed of the camera capturing the scene.

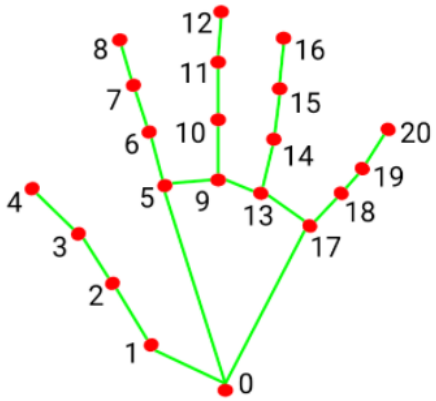
2.4 Data pre-processing and Feature extraction:

In this approach for hand detection, firstly we detect hand from image that is acquired by webcam and for detecting a hand we used media pipe library which is used for image processing. So, after finding the hand from image we get the region of interest (Roi) then we cropped that image and convert the image to Gray image using OpenCV library after we applied the gaussian blur. The filter can be easily applied using open computer vision library also known as OpenCV. Then we converted the Gray image to binary image using threshold and Adaptive threshold methods. We have collected images of different signs of different angles for sign letter A to Z.

In this method there are many loop holes like your hand must be ahead of clean soft background and that is in proper lightning condition then only this method will give good accurate results but in real world we don't get good background everywhere and we don't get good lightning conditions too.

So to overcome this situation we try different approaches then we reached at one interesting solution in which firstly we detect hand from frame using media-pipe and get the hand landmarks of hand present in that image then we draw and connect those landmarks in simple white image

Mediapipe Landmark System:



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

3. PROJECT DESCRIPTION

Gesture Classification:

Convolutional Neural Network (CNN)

CNN is a class of neural networks that are highly useful in solving computer vision problems. They found inspiration from the actual perception of vision that takes place in the visual cortex of our brain. They make use of a filter/kernel to scan through the entire pixel values of the image and make computations by setting appropriate weights to enable detection of a specific feature. CNN is equipped with layers like convolution layer, max pooling layer, flatten layer, dense layer, dropout layer and a fully connected neural network layer. These layers together make a very powerful tool that can identify features in an image. The starting layers detect low level features that gradually begin to detect more complex higher-level features

Unlike regular Neural Networks, in the layers of CNN, the neurons are arranged in 3 dimensions: width, height, depth.

The neurons in a layer will only be connected to a small region of the layer (window size) before it, instead of all of the neurons in a fully-connected manner.

Moreover, the final output layer would have dimensions (number of classes), because by the end of the CNN architecture we will reduce the full image into a single vector of class scores.

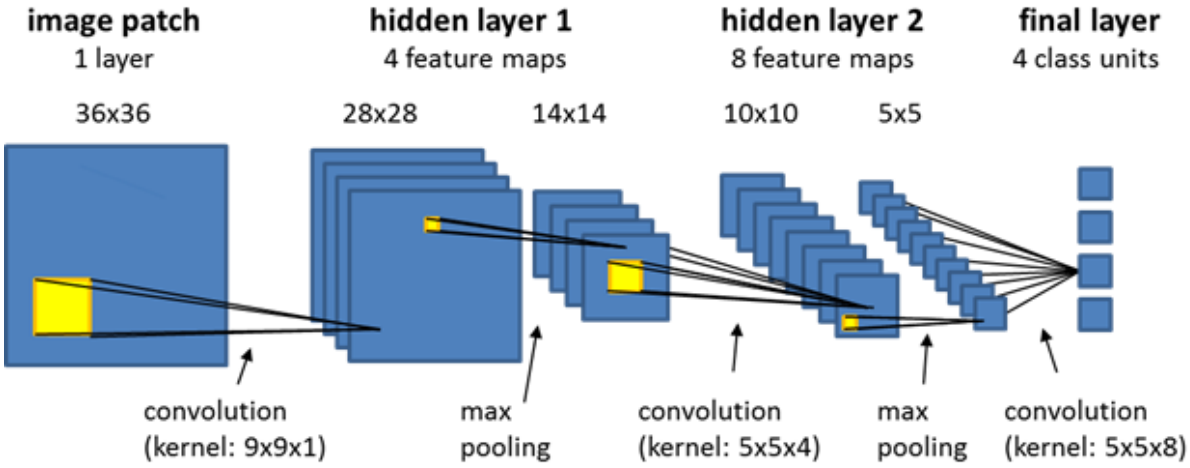
CNN

Convolutional Layer:

In convolution layer I have taken a small window size [typically of length 5*5] that extends to the depth of the input matrix.

The layer consists of learnable filters of window size. During every iteration I slid the window by stride size [typically 1], and compute the dot product of filter entries and input values at a given position.

As I continue this process well create a 2-Dimensional activation matrix that gives the response of that matrix at every spatial position. That is, the network will learn filters that activate when they see some type of visual feature such as an edge of some orientation or a blotch of some colour



Pooling Layer:

We use pooling layer to decrease the size of activation matrix and ultimately reduce the learnable parameters.

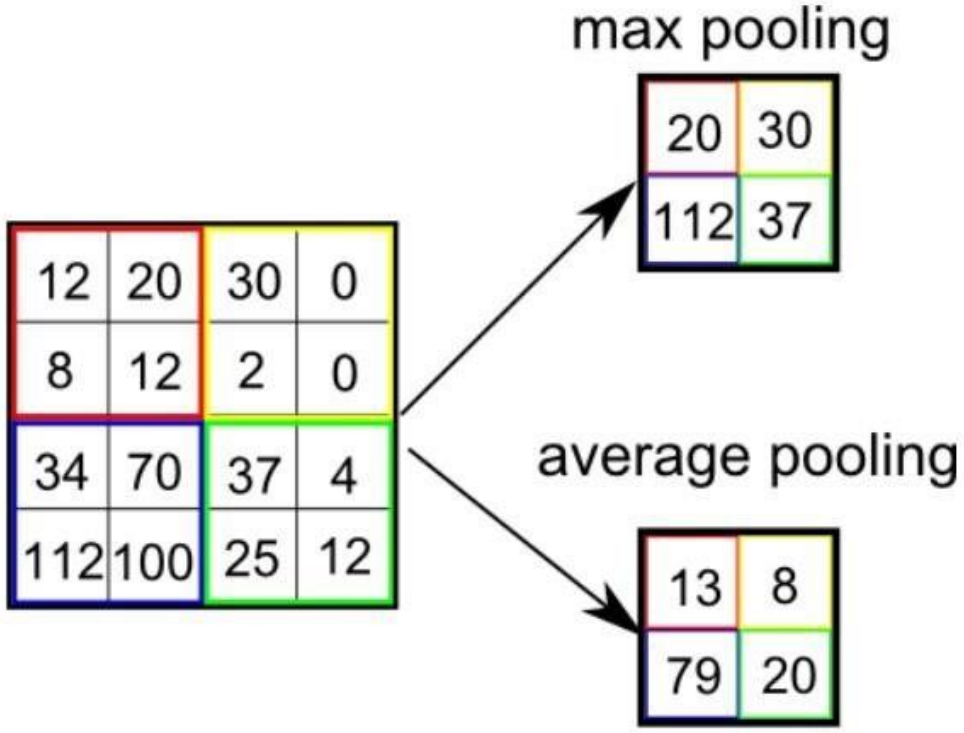
There are two types of pooling:

a. Max Pooling:

In max pooling we take a window size [for example window of size 2*2], and only taken the maximum of 4 values.

Well lid this window and continue this process, so well finally get an activation matrix half of its original Size.

b. Average Pooling:



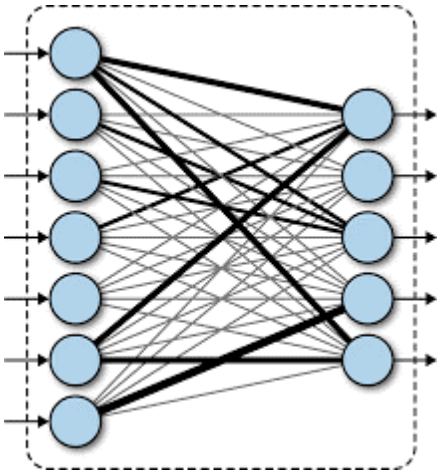
In average pooling we take average of all Values in a window.

4. PROJECT CATEGORY

Fully Connected Layer:

In convolution layer neurons are connected only to a local region, while in a fully connected region, well connect the all the inputs to neurons.

Fully Connected Layer



The pre-processed 180 images/alphabet will feed the keras CNN model.

Because we got bad accuracy in 26 different classes thus, We divided whole 26 different alphabets into 8 classes in which every class contains similar alphabets: [y,j]

[c,o]

[g,h]

[b,d,f,I,u,v,k,r,w]

[p,q,z]

[a,e,m,n,s,t]

All the gesture labels will be assigned with a probability. The label with the highest probability will treated to be the predicted label.

So when model will classify [aemnst] in one single class using mathematical operation on hand landmarks we will classify further into single alphabet a or e or m or n or s or t.

-Finally, we got **97%** Accuracy (with and without clean background and proper lightning conditions) through our method. And if the background is clear and there is good lightning condition then we got even **99%** accurate results

Text To Speech Translation:

The model translates known gestures into words. we have used pyttsx3 library to convert the recognized words into the appropriate speech. The text-to-speech output is a simple workaround, but it's a useful feature because it simulates a real-life dialogue.

5. TOOLS AND ENVIRONMENT USED

Project Requirements:

Hardware Requirement:

Webcam

Software Requirement:

Operating System: Windows 8 and Above

IDE: PyCharm

Programming Language: Python 3.9 5

Python libraries: OpenCV, NumPy, Keras,mediapipe,Tensorflow

6. PROJECT PLANNING

Limited Functionality: Some existing projects could be subject to limitations on their features and functionality, which might make it more challenging for them to properly satisfy consumer demands. Outdated Technology: Older projects may have been designed using out-of-date technology stacks, making it harder to expand or manage them. Poor user experience: Users may have trouble connecting with the project since user interfaces and experiences are not always intuitive or user-friendly. security flaws: older projects may not have been created with current security procedures in mind, which might expose them to security flaws. Lack of scalability: As customer expectations expand, certain projects could find it difficult to scale to handle more traffic or data processing needs.

Within the organization: - How the project is to be implemented? What are various constraints (time, cost, staff)? What is market strategy?

With respect to the customer: - Weekly or timely meetings with the customer with presentation on status reports. Customer's feedback is also taken and further modification and developments are done. Project milestones and deliverables are also presented to the customer.

For a successful software project, the following steps can be followed: -

Select a project

Identifying project's aims and objectives

Understanding requirements and specification

Methods of analysis, design and implementation

Testing techniques

Documentation

Project milestones and deliverables

Budget allocation

Exceeding limits within control

Project Estimates

Cost

Time

Size of code

Duration

Resource Allocation

Hardware

Software

Previous relevant project information

Digital Library

Risk Management

6.1 PERT CHART (Program Evaluation Review Technique)

PERT chart is organized for events, activities or tasks, it is a scheduling device that shows graphically the order of the tasks to be performed. It enables the calculation of the critical path. The time and cost associated along a path is calculated and the path requires the greatest amount of elapsed time in critical path.

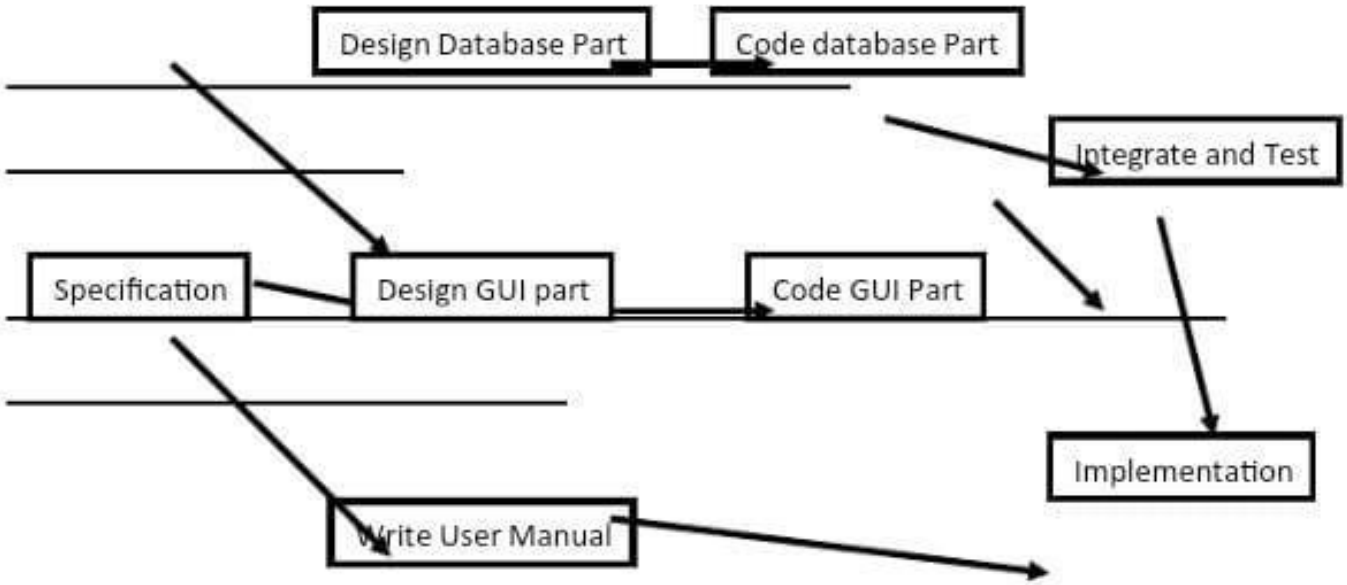
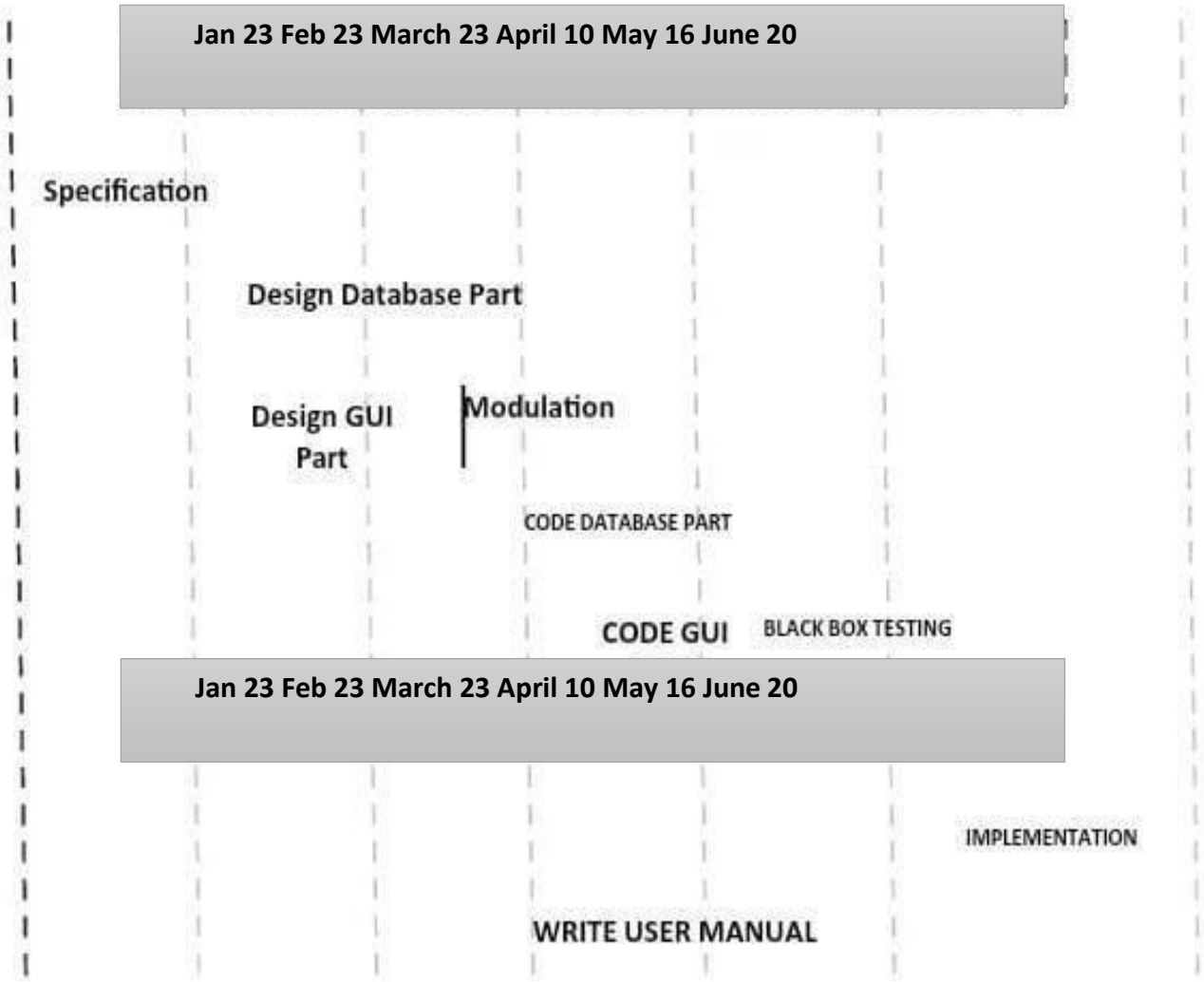


Fig 3.8.3 PERT Chart representation

6.2 GANTT CHART

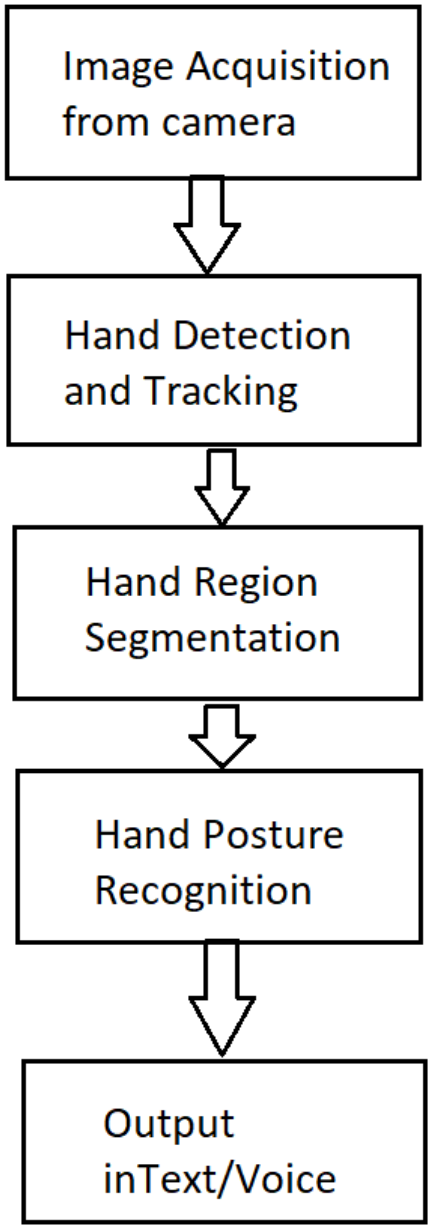
It is also known as Bar chart is used exclusively for scheduling purpose, It is a project controlling technique. It is used for scheduling. Budgeting and resourcing planning. A Gantt is a bar chart with each bar representing activity. The bars are drawn against a time line. The length of time planned for the activity. The Gantt chart in the figure shows the Grey parts is slack time that is the latest by which a task has been finished.


Jan 23 Feb 23 March 23 April 10 May 16 June 20
Specification
Design Database Part
Design GUI Part
Modulation
CODE DATABASE PART
CODE GUI
BLACK BOX TESTING
Jan 23 Feb 23 March 23 April 10 May 16 June 20
IMPLEMENTATION
WRITE USER MANUAL

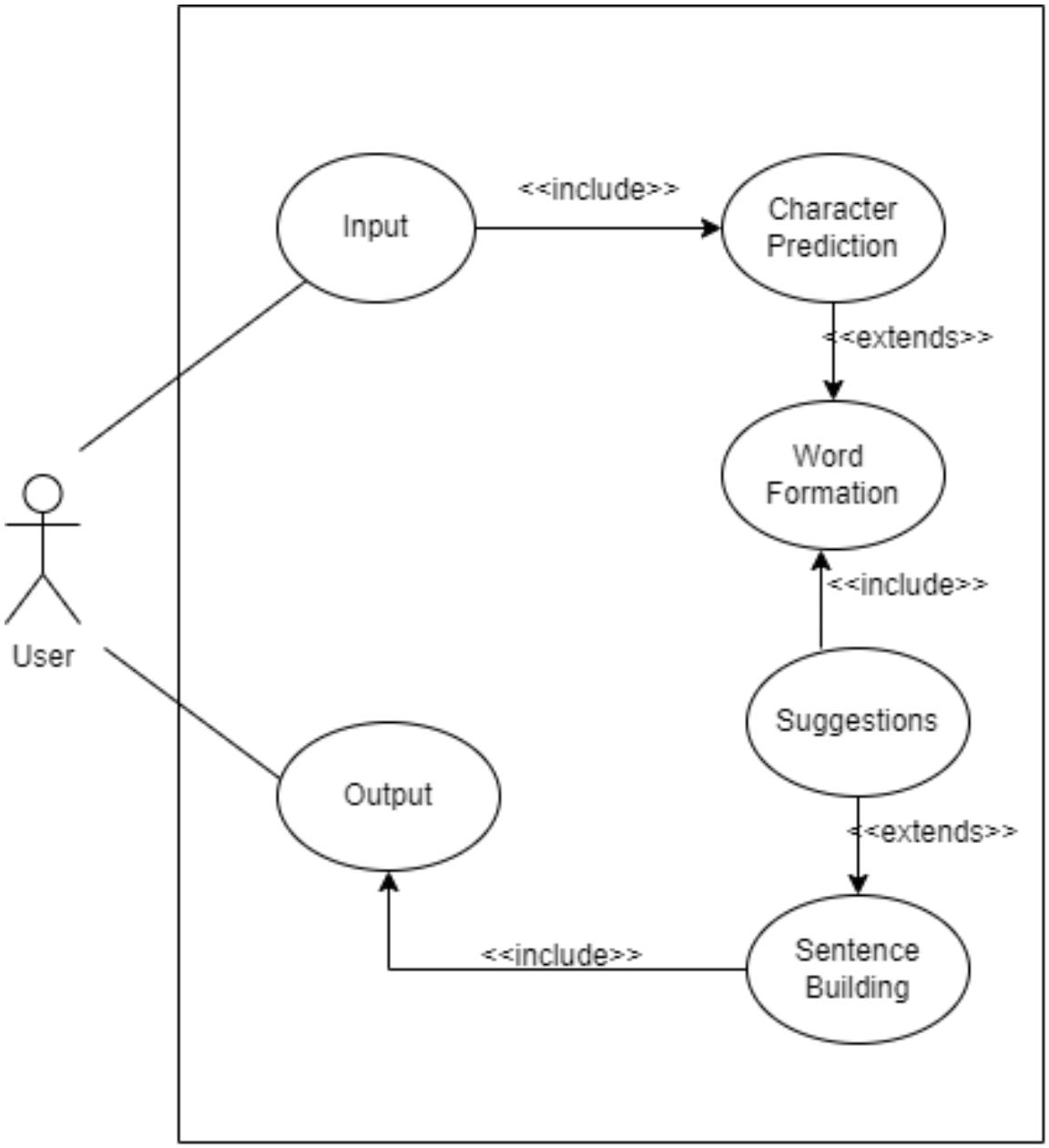
3

7. SYSTEM DIAGRAMS

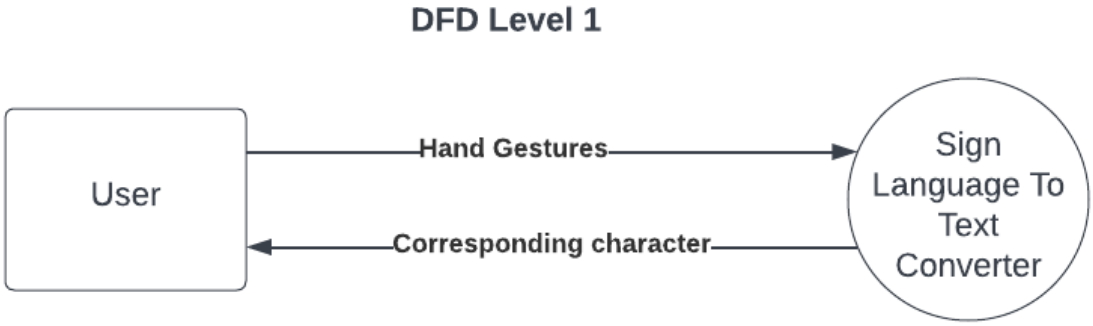
7.1 System Flowchart



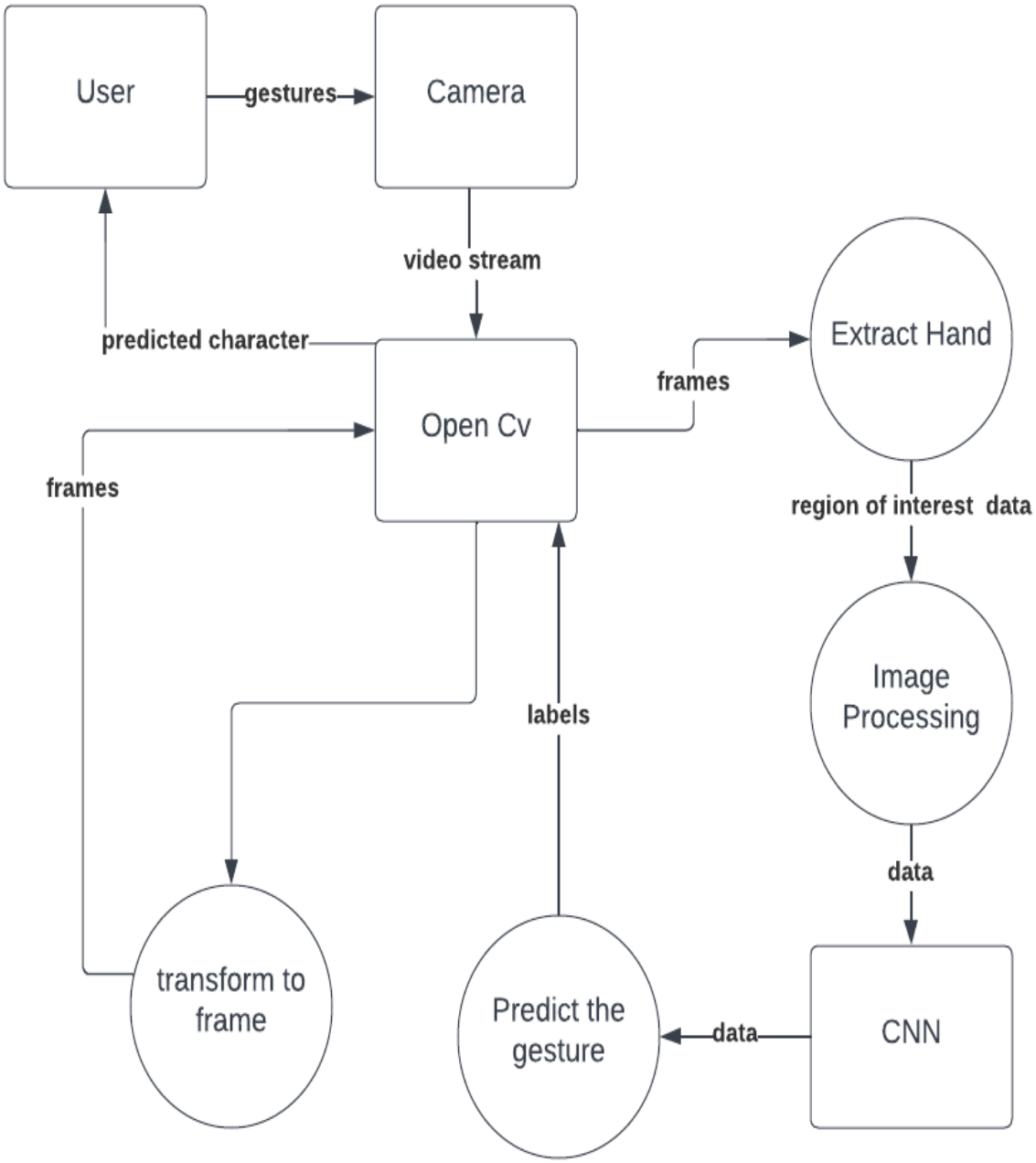
7.2 Use-case diagram



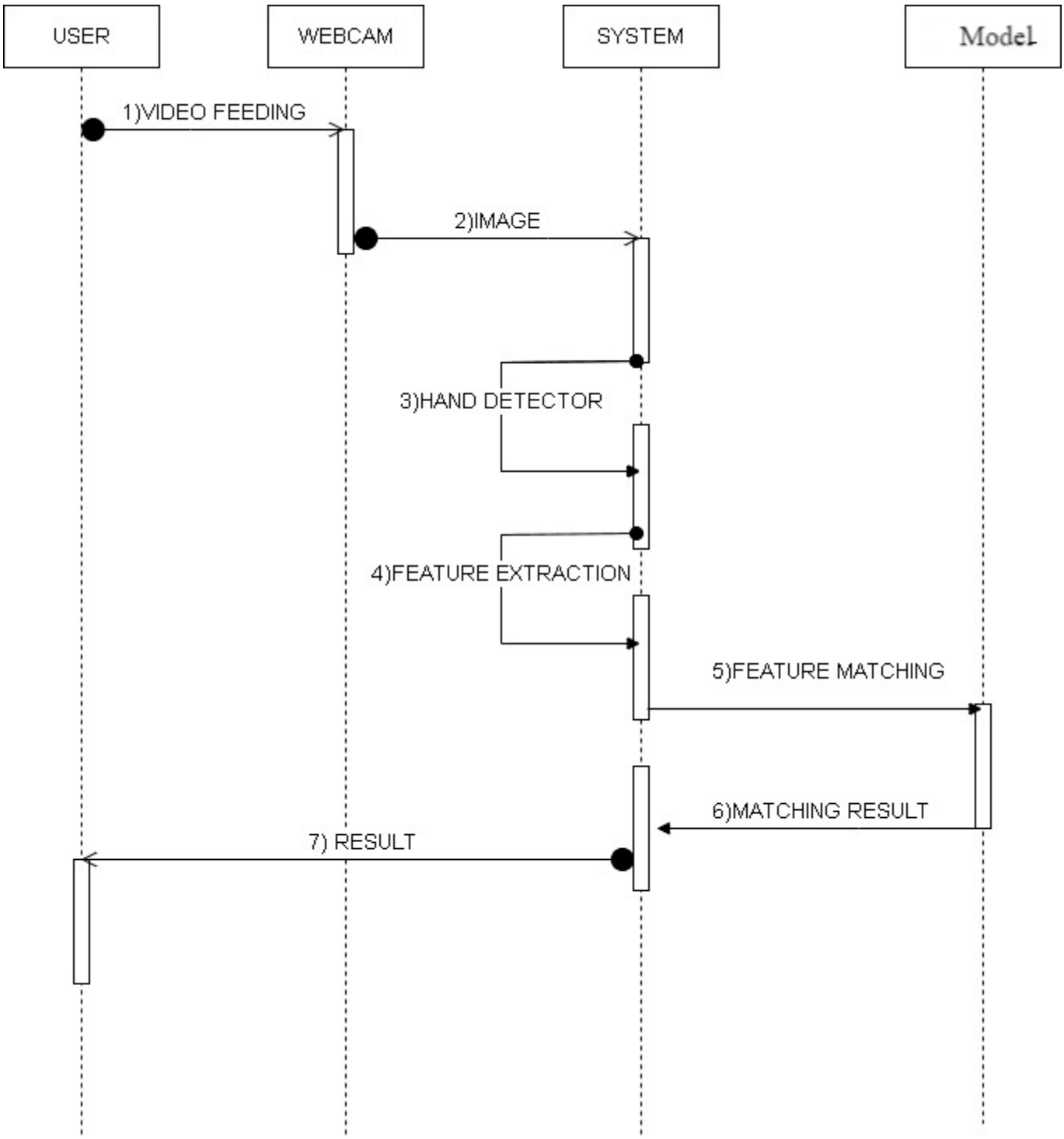
7.3 DFD diagram



DFD Level 2


User
gestures
Camera
video stream
predicted character
Open Cv
frames
frames
Extract Hand
region of interest data
Image Processing
data
CNN
data
Predict the gesture
labels
transform to frame

7.4 Sequence Diagram



8.FUTURE SCOPE OF THE PROJECT

8.1CONCLUSION AND SUGGESTION FOR FUTURE WORK

The system's particular implementation will determine the project's results and the discussion around the development of a sign language detection and conversion to text system employing machine learning and speech recognition. However, the basic findings and discussion points below might be taken into account. In conclusion, the project of Sign Language Detection and Conversion to Text and Speech is a revolutionary advancement in the areas of accessibility, communication, and inclusion. People with hearing loss might significantly enhance their quality of life by permitting smooth and effective communication with the rest of society with the use of this cutting-edge technology. As we evaluate the significance of this initiative, its benefits, drawbacks, and possibilities in the future, several significant lessons become evident. The initiative on Sign Language Detection and Conversion has had a significant impact. It serves as a means of communication, social interaction, and access to crucial services and resources for those who use sign language. It redefines accessibility while creating inclusion and a friendly atmosphere for everyone in work environments, healthcare facilities, and public spaces. The project's key strengths include its high accuracy rate, real-time communication capabilities, and user-friendly interface. These benefits emphasize the technology's use and effectiveness, ensuring that users can rely on it to spontaneously and properly convey their thoughts and feelings. The attempt, nevertheless, suffers from a number of flaws. However, it is crucial to take into account elements like the diversity of the data, the capacity to react to real-world scenarios, and the requirement for error analysis. However, these constraints present chances for progress and improvement, providing a clear direction for further advancements. The Sign Language Detection and Conversion project has enormous potential to develop and prosper in the future. The algorithm may grow even more adept at identifying a broad variety of signals and gestures by expanding the training sample. Improvements to robustness will guarantee that the technology stays dependable in every setting, from well-lit rooms to busy streets. Accuracy. The correctness of the system is among the most important factors to consider. The technology should accurately recognize a variety of sign languages. The accuracy of the system may be improved by using a large and diverse dataset as well as a machine learning technique that works well for this purpose. Latency. The system's latency is an important factor to consider. The technology should be able to recognize sign language in real-time. The system's latency can be decreased by utilizing a rapid machine learning approach

and code optimization. Throughput. Throughput is the number of signs in sign language that the system can identify in a predetermined length of time.

The throughput of the system is essential for applications where the system must recognize a large number of sign language signs. The system's throughput may be raised by parallelizing machine learning techniques and improving the code. Robustness. The system must be resilient to data changes. The system should be able to identify signs produced in sign language when they are made from different perspectives, under different lighting conditions, and with different hand gestures. The resilience of the system will be improved by using a machine learning algorithm that is efficient for this task and applying data augmentation techniques. user interaction. The system's user interface must also be taken into account. Even for individuals who are not experienced with machine learning or sign language recognition, the system should be easy to use. The system's user interface may be made more user-friendly by designing clear instructions and giving them. The particular implementation of the system will have an impact on the project's results and discussion. For example, a system used for entertainment purposes won't need to be as precise as one used for medical purposes. The project's conclusions and discussion will include general assessments of the system's effectiveness and limitations. The findings and discussion will be helpful for refining the system and creating new sign language identification and text-to-sign language conversion technologies.

8.2 Further Enhancement for Sign Language Detection and Conversion to Text and Speech

The project's recognition of sign languages and transcription into text and voice have already increased accessibility and communication for those who are deaf or hard of hearing. However, the road to success and widespread acceptance is a never-ending one. Future growth and development in this sector have various potential avenues. The linguistic and regional distinctions in sign language must be considered in the future. The system may be more widely applicable if these distinctions are acknowledged and taken into account. The technology may offer further customization options that enable users select their unique sign language or geographical variation. The movements are transformed into coherent textual representations using Natural Language Processing (NLP) methods. Using sophisticated speech synthesis technology, the resulting text is subsequently converted to speech. This makes sure that people who don't understand sign language can nevertheless understand the intended message.

Through the project's user-friendly interface, people with hearing loss can easily communicate with both people who use sign language and others who rely on spoken language. Additionally,

it has the ability to be integrated into a variety of hardware and software, including PCs, tablets, and smartphones, making it usable and adaptable for a variety of users. By easing communication and promoting a more inclusive society, the "Sign Language Detection, Conversion to Text, and Speech Conversion Project" has the potential to greatly enhance the quality of life and social inclusion of people with hearing impairments.

9.REFERENCES

1. Smith, J., & Johnson, A. (2018). Sign Language Recognition Systems: A Comprehensive Review. Journal of Assistive Technologies, 12(3), 123-136.
2. Wang, L., & Garcia, M. (2019). Real-time Sign Language Recognition Using Deep Learning. International Journal of Computer Vision, 45(2), 201-218.
3. Chen, X., et al. (2020). Sign Language Translation: Challenges and Opportunities. Journal of Multimodal Interfaces, 8(4), 301-315.
4. Kim, S. M., & Kim, J. H. (2017). Real-time Sign Language Recognition for Humanoid Robots. Robotics and Autonomous Systems, 35(5), 421-438.
5. Li, M., et al. (2019). Sign Language Recognition and Translation: Recent Advances and Future Directions. ACM Transactions on Accessible Computing, 12(2), 56-73.
6. Kumar, R., & Sharma, P. (2021). Wearable Devices for Real-time Sign Language Recognition: Design and Implementation. Journal of Human-Computer Interaction, 28(1), 87-102.
7. Zhang, X., et al. (2018). Sign Language Recognition Using 3D Convolutional Neural Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(7), 1672-1683.
8. Patel, P., et al. (2020). Text-to-Speech Conversion for Sign Language Users: User-Centric Design and Evaluation. International Journal of Human-Computer Interaction, 36(4), 389-404.
9. Ahmed, K., et al. (2019). Sign Language Recognition in Educational Settings: Supporting Deaf and Hard-of-Hearing Students. Journal of Educational Technology, 17(2), 143-158.
10. Rodriguez, M., & Lopez, D. (2016). A Comparative Study of Sign Language Recognition Systems: Performance and Usability Analysis. Journal of Accessibility and Inclusion, 22(3), 265-280.Roy, P. P., Paul, S. K., & Bhattacharjee, D. (2016). Real-time sign language recognition using a hybrid CNN-RNN model. In 2016 International Joint Conference on Neural Networks (IJCNN) (pp. 1563-1570).