MINOR-2 PROJECT

SYNOPSIS

For

Real-Time Sign Language Interpretation using Deep Learning

Submitted By:

Specialization	SAP ID	Name
CSE [Artificial Intelligence and Machine Learning]	500082307	Anuj Singh Naruka
CSE [Artificial Intelligence and Machine Learning]	500082531	Bhavy Kharbanda
CSE [Artificial Intelligence and Machine Learning]	500083965	Dhruv Gupta
CSE [Artificial Intelligence and Machine Learning]	500082411	Devashish Agarwal

Department of Informatics

School Of Computer Science

UNIVERSITY OF PETROLEUM & ENERGY STUDIES,

DEHRADUN- 248007. Uttarakhand

Dr. Shaurya Gupta
Project Guide

Dr. Thipendra P Singh
Cluster Head

School of Computer Science
University of Petroleum & Energy Studies, Dehradun

Synopsis Report

Project Title: - Real-Time Sign Language Interpretation using Deep Learning

ABSTRACT

Speech impairment constitutes a challenge to an individual's ability to communicate effectively through speech and hearing. To overcome this, affected individuals' resort to alternative modes of communication, such as sign language. Despite the increasing prevalence of sign language, there still exists a hindrance for non-sign language speakers to effectively communicate with individuals who primarily use sign language for communication purposes. Sign languages are a class of languages that employ a specific set of hand gestures, movements, and postures to convey messages. Deaf and Hard-of-Hearing (DHH) people require some type of technology to convey their feelings or if I rephrase it, we are the people required to learn sign language to be able to understand what a DHH person is trying to convey, and with a motive to convey our concern towards the same, we are here with our project.

The main motive behind this project is to provide such users with a user-friendly interface to convey their messages to a person who doesn't understand sign language using deep learning technology. The objective of this project is to conduct an analysis and recognition of various alphabets present in a database of sign images in real-time using the camera of the device to form sentences. This project will include the formation of a dataset on which such a model can be trained, that is we will be creating our dataset and won't be using any existing one. Further, using Convolutional Neural nets will be used to train a classifier to predict the alphabet gestured by the user. As a further extension to the project, our interface will also include a feature to show exactly the opposite of the same. It will be able to generate a sign sequence on the basis of an input sentence in order to educate people about using sign language. This function will help the user to communicate with the DHH person using their own method i.e., Sign language.

Keywords – Sign Language Translation, Speech impairment, Gesture recognition, Convolutional Neural nets, Deep learning, Deaf and Hard-of-Hearing (DHH)

INTRODUCTION

According to the World Health Organization (WHO), it is estimated that about 466 million people globally have some form of disabling hearing loss. This represents about 5% of the world's population. In terms of speech problems, the prevalence is harder to estimate, as it can vary greatly depending on the type and severity of the speech disorder, as well as access to healthcare and support services. However, speech disorders are known to affect a significant portion of the population, particularly children, and can have a profound impact on communication and quality of life.

Sign language is a form of visual-spatial communication utilized by a large number of individuals who are either deaf or hard of hearing. These languages have been specifically devised to cater to the needs of individuals who are deaf or have difficulty speaking. They utilize a precisely coordinated and synchronized sequence of hand movements, hand orientations, and hand shapes. It is believed that there are over 200 distinct sign languages utilized globally, with the understanding that even if their corresponding spoken languages are alike, the sign languages themselves may not necessarily be mutually understandable.

There are various sign languages prevalent in different regions, such as American Sign Language and Indian Sign Language. All these languages prefer a particular hand position but majorly, it is divided into two different categories, two-handed gestures, and one-handed gestures. American Sign language (ASL) is focused on one-handed gestures while Indian Sign Language (ISL) works on two-handed gestures. For this project, we will majorly, work on ASL as it is more spread and used by millions of people worldwide.

This project helps in the creation of a human-machine interface that can improve communication between healthy individuals and those with hearing impairments is a crucial issue, aimed at replacing the need for a human translator as the third party in the communication. **Real-Time Sign Language Interpretation** allows us to implement the same on a device that will detect these gestures and interpret them and will give the output alphabet related to that gesture. Our interface will help the people to form complete sentences so that the other person with no knowledge of sign language can understand the expression of the person with hearing and speaking disorders. Further, this interface will also allow the user to generate a sequence of signs by entering an input sentence sequence. We are implementing this feature to make it more comfortable to communicate even when don’t know about Sign language. They can simply imitate the generated sequence to convey their messages to the other person. The technicalities of this project are briefly explained in the methodology section.

Another extension to the project is regarding the choice of Sign language, as here we are focusing more on American Sign language as of now but if the time and deadlines allow us, we will implement the same model on Indian Sign language as well. So, the user would be able to choose even the language on our interface and make it more user-friendly and convenient to use.

LITERATURE REVIEW

We referred a several literatures to find the inspiration and in-depth knowledge of the subject. After going through various research papers and journals the insights that we integrated together are as follows:

Sequence modelling has various uses, including machine or language translation. In particular, neural machine translation involves converting a text from one language to another. [1] Here, Sign language is also a type of language so it implies that sign language conversion also borrows concepts from various natural language processing (NLP) tasks where sentences from one language are translated to the other. Recognizing gestures and actions falls under the umbrella of sequence modelling. It's commonly viewed as a problem of categorizing. The models in these cases focus largely on the temporal features of a video, particularly the movements of the body, hands, and face. CNN models are used for the purpose of image manipulation or capturing the frames of a video for the same. Importance of Computer vision in our project was explained in this paper. [3]

We also needed some basic background check on the existing technologies and models which were related to the project we are going to develop. So, for that we referred this paper to get our facts clear. [2] It also showed the results in the form of a comparative analysis on the use of various Deep learning models for the same. So, it helped us to take a wise choice for the use of model. [4]

Skin segmentation or the colour of skin while training the model can bring drastic changes to the performance of the mode. This research paper gave us the idea to create our own dataset using two different persons with different skin colours. This will give us a better chance to increase the accuracy of the model as the training on such dataset will be better. [5]

Existing models on Indian Sign Language recognition are very less in number as the language itself requires the motion of hands which is very difficult to capture in an image frame. LSTM models with attention layers and seq2seq models are used to rectify the problem and to propose a better solution as they have some memory to store the result and can decide what to forget before passing it forward to the next layer. [6]

PROBLEM STATEMENT

The problem statement is very simple in our case, we want to provide a tool for people with hearing and speaking impairments that can help them to convey their needs, feelings, and emotions through alphabets and sentences. This project is personally very close to me as my aunt suffers from a Deaf and Hard of hearing problem. So, during childhood, it was very difficult for me to understand her, and it made me feel miserable so that’s another reason we are developing this project. I always wanted to do something for her as I might learn the language but not everyone else can. So, we require a tool to translate the signs into an alphabet.

It is unfortunate that the majority of individuals with normal hearing abilities do not possess an understanding of sign language, leading to potential missed opportunities for members of the Deaf and Hard-of-Hearing (DHH) communities in various spheres of life such as education, employment, sports, and other social activities. Sign language interpretation can mitigate these challenges by promoting communication between sign language users and non-users. However, the utilization of human sign language interpreters can prove to be inconvenient, challenging to arrange, and expensive. To address this issue, several research groups are exploring the possibility of automating the translation process through the application of machine learning techniques and so are we. It is crucial to acknowledge that artificial intelligence tools, such as the one being proposed, serve to provide language support to both DHH sign language users and non-users. All of these scenarios give us the foundation problem for our project. That's why we are here with this project to rectify the problem.

OBJECTIVES

The main inspiration/objective behind the idea was the struggle that we saw among the people suffering from hearing and speaking detriments. The primary objective of this project is to utilize deep learning technology to create a user-friendly interface, enabling individuals who use sign language to effectively communicate with those who do not understand it. Our main aim is to improve accessibility and communication for individuals who are deaf or hard of hearing. The use of deep learning technology in sign language interpretation has the potential to provide these individuals with a more effective and efficient means of communicating with the wider community, enabling them to more fully participate in society and overcome barriers to communication. We have observed different research groups running in various cities and countries, to provide easy and effective solutions for the same problem, So, why not make one by ourselves? Sub-objectives related to this project are as follows:

- **Bridging the Communication Gap:** Sign language translators using deep learning can help bridge the communication gap between individuals who are deaf or hard of hearing and the wider community, enabling them to more fully participate in society.
- **Ease of Use:** A sign language translator using deep learning is expected to be more user-friendly and easier to use than traditional human sign language interpreters, making it easier for individuals who are deaf or hard of hearing to communicate with others.
- **Cost-effectiveness:** Human sign language interpreters can be costly, especially for individuals who need to use their services regularly.
- **Improved Accuracy:** Deep learning algorithms can learn from a large number of examples, making them better at recognizing patterns and improving their accuracy over time. This can lead to improved accuracy in sign language translation compared to traditional methods.

METHODOLOGY

We have divided the methodology into five phases -

Dataset Creation – As our project is focused on sign detection the dataset will be generated by different signs generated by us using our hands. Now there are numerous variations in each sign being generated in terms of skin colour, size of hand, angle with which it has been captured, clarity of it etc. Labels for each attribute is created in the dataset which is then further augmented and stored in our local machine. Labels that are created in the dataset play an integral role for the entire project as they form the basis which are taken into consideration when the data is passed through the model.

Pre-Processing – This phase of the project deals with simplifying the dataset that has been created. Here a lot of unnecessary elements are segregated and removed like background, different noises present in the image etc. leaving us with the areas of interest which in turn is the required information. Another integral part of pre-processing is image binarization. As machine understands and processed binary data the images are converted into binary format having different specification for each pixel, grey area etc. Each configuration of 1's and 0's carries integral information regarding the dataset. Some other steps of image processing are also completed here like smoothing of images, creating threshold and gradients of each image. After this the pre-processed data is now ready to be passed through the model.

Model – For processing these images we'll be using CNN or convolutional neural network model. CNN consists of three types of layers that are convolutional layer, pooling layer and fully connected layer. Each layer performs different task on the rendered image that we have provided after pre-processing the dataset. For example, the pooling layer works on the dimensionality of the feature map that is created after the image passes through the convolutional layer. As each unit of dataset passes through this model, it learns and extracts common features which eventually increase its processing speed. It will classify the image frame and will return the appropriate match of alphabet to the user. The model will use an NLP model as well to predict the words that can be created using those alphabets and will help in completing the sentences of the user. The output generated by this trained model has to be displayed and that's where the work of frontend kicks in.

Frontend – This phase is all about displaying the output generated by the model. We have used OpenCV and Flask for displaying the output thus making it more interactive and easier for the user to understand. Flask is a tool associated with python for development and OpenCV is a well-known commodity in terms of graphics and image generation. This interface will provide an easy way out to visualize the words and sentences.

Testing ad Improvement – After the model has been trained on the dataset which was provided to it, we test it on some other dataset thus dealing with different metrics of accuracy, precision etc.

METHODOLOGY – DATASET CREATION AND PRE-PROCESSING

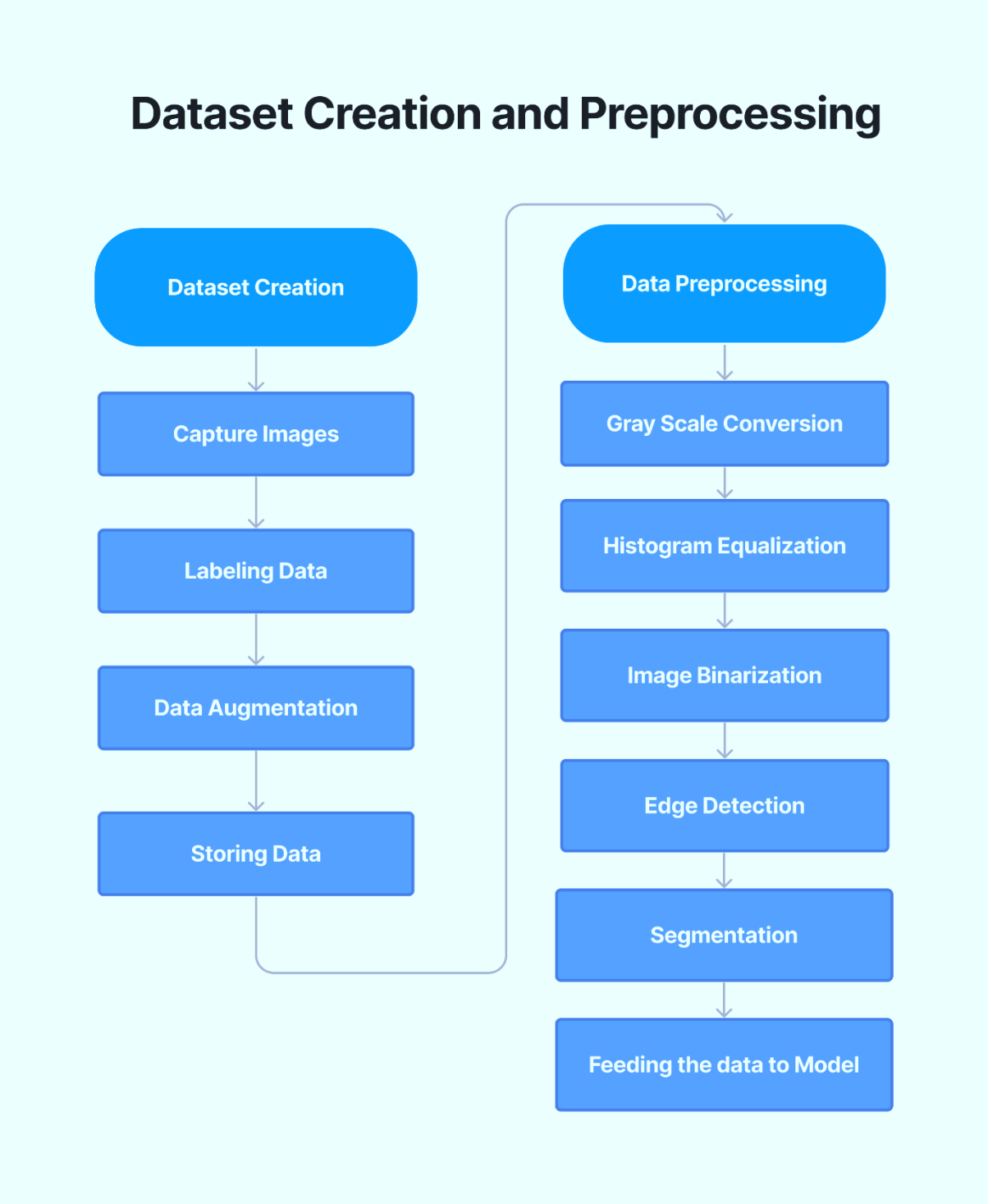

Figure 1: Dataset Creation and Image Pre-Processing Flowchart

METHODOLOGY - WORK FLOW

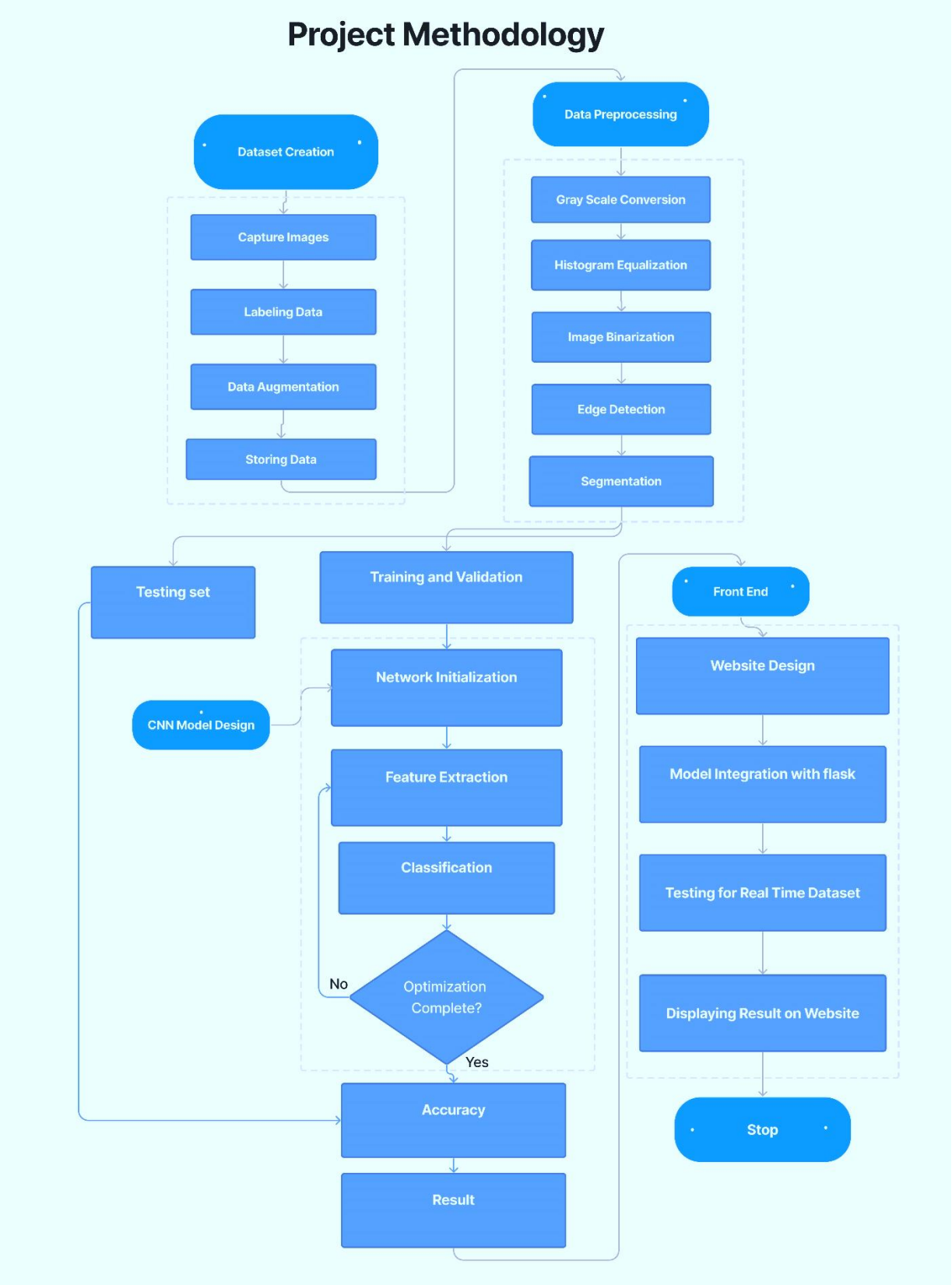

Figure 2: Project Methodology

Talking about the SEPM models, we will be requiring constant feedback and that's how we will be updating our model. On the back of our mind, we will try to use the RAD approach because we need the rapid development and it would be a perfect approach for a shorter deadline and recursive feedback. It requires short time span, requirements are already known to us, technical risk is very limited in our scenario, and the model will be flexible for change.

UML Diagram:

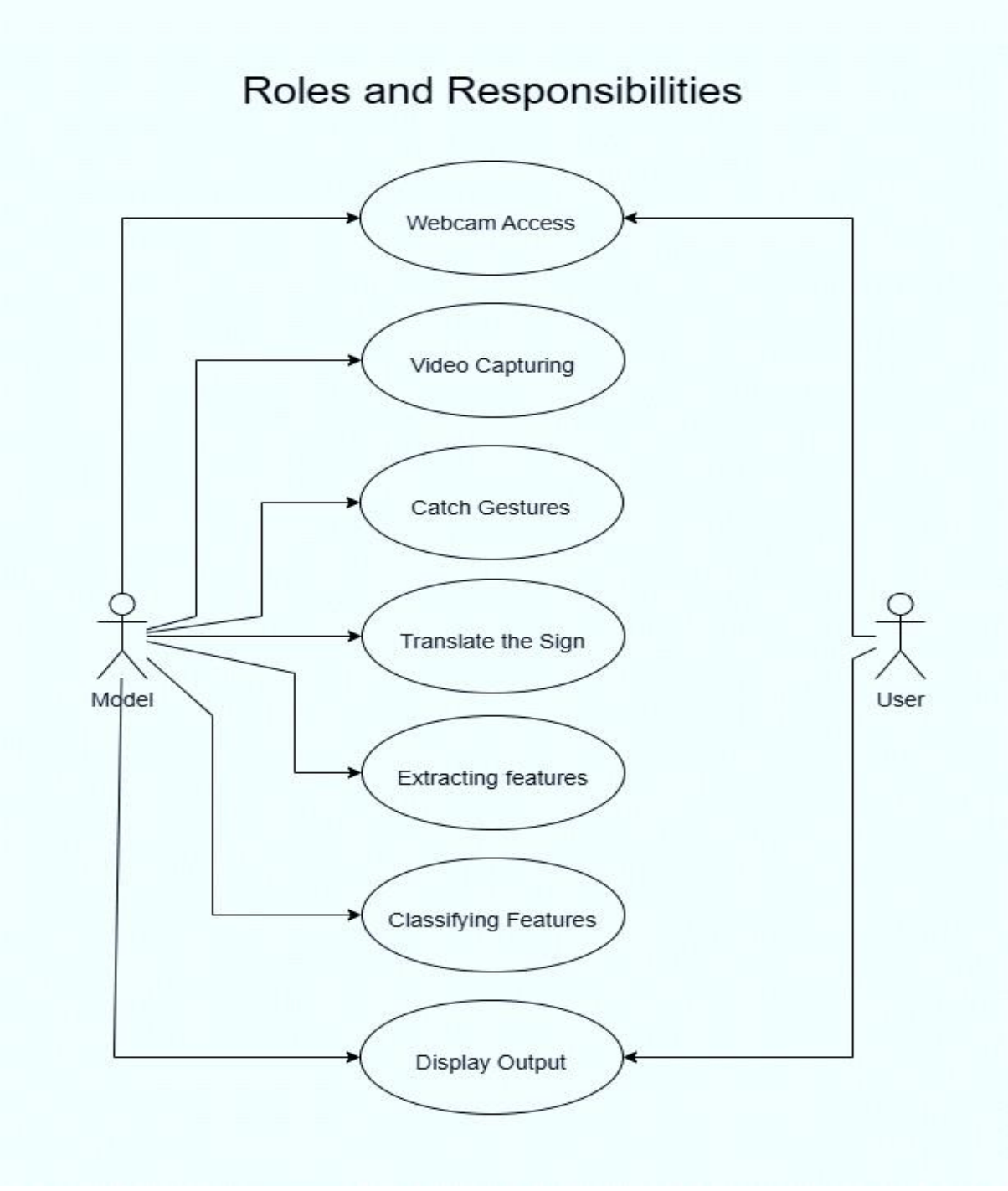

Figure 3: UML Diagram for Roles and Responsibilities

PERT CHART

Pert Chart is a project management tool used to schedule, organize, and synchronize various phases of the project.

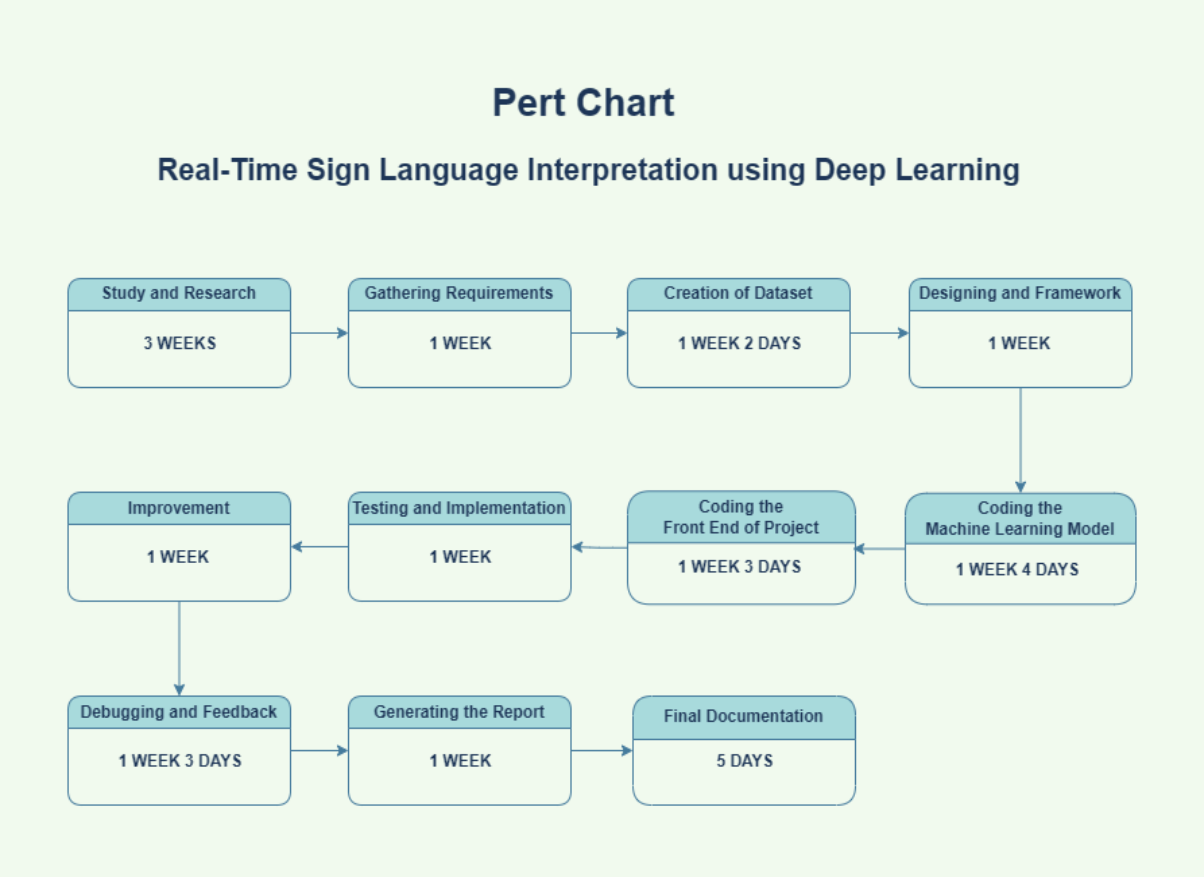

Figure 4: PERT Chart for the project

GITHUB LINK

The GitHub link to the repository of our project:

https://github.com/KharbandaBhavy/Real-Time-Sign-Language-Interpretation-usign-Deep-Learning

REFERENCES

[1] Tejaswini Ananthanarayana, Priyanshu Srivastava, Akash Chintha, Akhil Santha, Brian Landy, Joseph Panaro, Andre Webster, Nikunj Kotecha, Shagan Sah, Thomastine Sarchet, Raymond Ptucha, and Ifeoma Nwogu. 2021. "*Deep Learning Methods for Sign Language Translation.*" ACM Trans. Access. Comput. 14, 4, Article 22 (December 2021), 30 pages. Available: https://dl.acm.org/doi/10.1145/3477498

[2] Abiyev, Rahib H *"Sign Language Translation Using Deep Convolutional Neural Networks".* Vol. 14, KSII Transactions on Internet and Information Systems. Korean Society for Internet Information (KSII); 2020. Available: https://koreascience.kr/article/JAKO202011161036129.page

[3] K. Bantupalli and Y. Xie, *"American Sign Language Recognition using Deep Learning and Computer Vision,"* 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4896-4899, doi: 10.1109/BigData.2018.8622141. Available: https://ieeexplore.ieee.org/document/8622141

[4] S. He, *"Research of a Sign Language Translation System Based on Deep Learning,"* 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), Dublin, Ireland, 2019, pp. 392-396, doi: 10.1109/AIAM48774.2019.00083. Available: https://ieeexplore.ieee.org/abstract/document/8950864

[5] S. Shahriar et al., *"Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning,"* TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 1168-1171, doi: 10.1109/TENCON.2018.8650524. Available: https://ieeexplore.ieee.org/document/8650524

[6] P. Likhar and R. G. N, *"Indian Sign Language Translation using Deep Learning,"* 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC), Bangalore, India, 2021, pp. 1-4, doi: 10.1109/R10-HTC53172.2021.9641599. Available: https://ieeexplore.ieee.org/abstract/document/9641599