National University of Computer and Emerging Sciences

GDS (Gesture Decode System)

FYP Team

Asma Ahmed……19L-2251
Raida Munir……19L-1162
Faseeh Ullah Jafar……19L-2273

Supervised by

Dr. Saira Karim

FAST School of Computing

National University of Computer and Emerging Sciences

Lahore, Pakistan

May 2023

Anti-Plagiarism Declaration

This is to declare that the above publication produced under the:

Title: GDS (Gesture Decode System)

is the sole contribution of the author(s) and no part hereof has been reproduced on **as it is** basis (cut and paste) which can be considered as **Plagiarism**. All referenced parts have been used to argue the idea and have been cited properly. I/We will be responsible and liable for any consequence if violation of this declaration is determined.

Date: May 25, 2023

Student 1

Name: Asma Ahmed
Signature:

Student 2

Name: Raida Munir
Signature:

Student 3

Name: Faseeh Ullah Jafar
Signature:

Authors' Declaration

This states Authors' declaration that the work presented in the report is their own, and has not been submitted/presented previously to any other institution or organization.

Abstract

The primary objective of this project is to empower individuals who are unable to speak by providing them with a virtual voice, facilitating easier communication. To achieve this, the proposed system will leverage deep learning mechanisms to convert Urdu sign language into Urdu text. This will be accomplished by extracting landmarks from the upper half of the body. The project encompasses various medical terms related to sign language, enabling comprehension of human body parts and diseases.

To build a comprehensive dataset for Urdu sign language, data will be collected through multiple methods. These include gathering information from hearing-impaired students and utilizing the resources available on the Pakistan Sign Language (PSL) website. By incorporating diverse sources, the system aims to ensure accuracy and inclusivity.

Furthermore, the system will operate in real time, functioning as a detection module. It will capture a person's miming gestures and identify the corresponding gesticulations, ultimately converting them into a coherent narrative. This real-time capability enhances the system's usability and efficiency, enabling seamless communication for individuals who rely on sign language as their primary means of expression.

Executive Summary

The following proposal ‘Gesture Decode System’ consists of the adaptation of Pakistani Sign Language whose ultimate goal is to output the essential words for the special people. In the modernized era where everything is going too fast, special people need to keep a translator and need extra struggle to reach at the equal position.

This project is addressing the need of sign language translator and providing a mechanism to render the Pakistani Sign Language. It is a system which will capture the person’s gestures with camera and identify the landmarks for the upper half body by extracting features and will label them with their enumerated names. It is mainly covering some of the medical terms specified in goals and objectives. This process will continue to search the specified gestures in the dataset and display the text on the screen.

The major limitation for this project is the availability of the dataset to train which will be collected using different methods. In contrary to Pakistani Sign language, there are numerous software to convert different sign languages whose dataset are extensively found on different websites.

This research could lead to future projects involving more extensive software for continuous Pakistani sign language recognition and translation and cover more domains. Moreover, this project could be the initiative to help the hearing-impair and dumb people of Pakistan admitting equally.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Communication gap is the huge barrier between mute and normal people. Mute people are those who have no or less speaking ability. They use sign language for conveying their message. Every region or country has their own sign language. Similarly, in Pakistan, PSL (Pakistan Sign Language) is being used. But, this is problematic for normal people to understand sign language.

The solution would be a digital decoder which decodes sign language to normal language. We will make a website for Pakistani Sign Language translation which will help those people by providing a virtual tongue to them and converts it into text by one video shot. We will use deep learning appliances to train our model on the people's gestures so that it can predict similar gestures.

Our project structure includes different sections. Section 1.1 will discuss the purpose of the document, section 1.2 is about the intended audience whereas section 1.3 will contain the definitions, acronyms and abbreviations of the article. Moreover, chapter 2 will accommodate the project vision, domain, goals, objectives and the scope of the method by briefly explaining the problem catered in this project. In addition to this, chapter 3 is about the detailed literature review of the related works. Furthermore, chapter 4 is describing software and hardware requirements of the model by briefly explaining the use cases, its graphical user interface and database design. Whereas, proposed approach has been discussed in chapter 5. Besides, chapter 6 is about the high level and low level design of the model. In addition to this, implementation and experimental results have been discussed in chapter 7 and 8 respectively. Moreover, chapter 9 will discuss about the user manual of our system. Lastly, chapter 10 is concluding the report by giving the recommendations for future work.

1.1 Purpose of this Document

The purpose of this research is to initiate the work on Pakistani Sign language and provide an effective and efficient way to facilitate the hearing-impair people to communicate easily. Gesture Decode system helps in modeling the system so that the special people would not have to face any inadequate consequences due to their disability.

1.2 Intended Audience

The main intended audience of the research article are the people who have to propose the similar system of translating sign language into text. This project is not only providing the methodology but the challenges that can be faced.

In addition to this, concerned people who have to use this software will be able to comprehend the software by reviewing the certain sections.

1.3 Definitions, Acronyms, and Abbreviations

SDG: Sustainable Development Goal

GDS: Gesture Decode System

Special people: Dumb or mute people (who cannot speak)

PSL: Pakistan Sign Language

1.4 Conclusion

Gesture Decode System helps to translate gestures into text. In this research, we have mentioned the motive to provide the virtual tongue to the hearing-impair people. Chapter 1 is giving a brief overview of our project. Furthermore, we have explained the goal of GDS which is to provide a good communication source to those people who need some instrument or an instructor for successful communication and discussed related work in the upcoming chapters.

Chapter 2: Project Vision

The problem catered in gesture decode system is to disperse the connection differences among doctors and hearing-impair patients with the help of deep learning techniques and landmarks identification with media-pipe. It mainly covers some of the human diseases and anatomy in Urdu. The ultimate goal of the proposed system is to facilitate and expedite the dumb people of Pakistan to actively participate in every field and treated equally with any umbrage of their disability.

2.1 Problem Domain Overview

This system will terminate or lessen the language barrier between tongue-tied and the normal people. The project's major objective is to help the unspeakable Urdu language patients to share their problems easily with the doctors and get their treatment done within the normal budget by decoding their hand movements.

2.2 Problem Statement

Unlike Pakistani sign language, there are many related works has been done on sign language translation in Arabic, English, Chinese, Hindi and German. The presented research is tackling with the problem of translating Urdu sign language to help Hearing-impair people to participate in every activity or field without extra effort and get education with normal fees.

2.3 Problem Elaboration

Sign language users are often underprivileged of functional communication and feel left alone because of rare social interactions. As discussed earlier, there is no previous software or system which helps to convert the Pakistani Sign language. The main reason could be the deficiency of dataset and the availability of translating mechanisms. The motivation of this paper is to fill that transmission gap between them by using assistive technology that allows the dumb people to communicate in their own language.

2.4 Goals and Objectives

The main objective of the project is to grant an easy communication source to dumb people by translating their sign language to the understandable language of their region. For this purpose, our system will input their gestures and output the translated language in the form of text. We are covering 25 human body anatomy present on PSL.

- People can get their check-ups done from normal doctors and would not be needing to search for special doctors who understand their language.
- This project is also cost efficient as they would not be needing to pay extra to the doctors who can understand their language.
- Our system is also helpful for the students who want to learn medical science.

2.5 Project Scope

Gesture Decode System mainly uses deep learning mechanisms to extract the image features and convert them into text. Our project consists of the following two modules:

- Media pipe holistic to extract the landmarks from the person's gestures by converting the video into image frames.

- LSTM architecture has been used to train the model.

2.6 Sustainable Development Goal (SDG)

People who have any disability are not considered normal such as the hearing-impair individuals have the deficiency of speaking and hearing. They face many difficulties in getting education and doing other daily life activities. They need more proficiencies and struggle than the normal people to achieve their target or goals. The targeted SDG is to reduce the inequality of special people and treating them as normal mankind. For this purpose, gesture decode system helps them to participate in daily life activities by providing them a translator of their sign language.

Figure 1: Reduced Inequalities
This figure represents the targeted SDG of our FYP (Reduced Inequalities)

2.7 Constraints

The methodology and objectives to achieve the specified goal is managed and defined properly. But following are the major constraints to be faced in this project:

- First experimental try of translating Urdu Sign language and thus no algorithm is available for Urdu segmentation.
- No developed dataset available for Pakistani sign Language.

2.8 Business Opportunity

This project is covering the specified scope of Pakistani Sign Language translation of some medical terms. It is the initiative work in generating Urdu sign language corpus and its translation training dataset. This concept might be developed into a company to provide a full-fledged real-time continuous Urdu sign language detection and translation service to help the hearing-impair people bridge their communication gap.

2.9 Conclusion

In this chapter we have briefly described the project vision of GDS (Gesture Decode System), elaborated the scope, goals and objectives. Furthermore, the constraints, business opportunity and Sustainable Development Goal has been discussed.

Chapter 3: Literature Review / Related Work

Literature review plays a crucial role to clearly identify the scope and gap of any project. In this research, we have studied almost all the related articles for the last three years from Google scholars and proposed the summary of every article in the specified section. In the summary section, we have discussed the purpose and the method used in particular paper. Then, we have talked about the strengths and limitations of every article and finally its relationship with our research.

3.1 Definitions, Acronyms, and Abbreviations

CNN: Convolution Neural Network

RNN: Recurrent Neural Network

ML: Machine Learning

CV: Computer Vision

OpenCV: An open-source library for ML and CV techniques.

ReLU: Rectified Linear Unit

NMT: Neural Machine Translation

CSLR: Continuous Sig Language Recognition

SLT: Sign Language Translation

LSTM: Long Short-Term Memory Networks

HMM: Hidden Markov Model

GRU: Gated Recurrent Units

WER: Word Error Rate

ISL: Indian Sign Language

mLTSF: Multi-scale Local-Temporal Similarity Fusion

ResNet: Residual neural Network

SVM: Support Vector Machines

PTC: Position-Aware Temporal Convolver

CFS: Content-aware Feature Selector

GSL: Greek Sign Language

CSL: Chinese Sign Language

SL: Sign Language

GAN: Generative Adversarial Network

SI: Signer Independent

SKE: Skeleton

CTC: Connectionist Temporal Classification

VRML: Virtual Reality Modeling Language

MOPGRU: Media-Pipe Optimized Gated Recurrent Unit

HSV: Hue Saturation value

DTW: Dynamic Time Wrapping

CRF: Conditional Random Field

S2G: Sign to Gloss

S2G2T: Sign to Gloss to Text

G2T: Gloss to Text

3.2 Detailed Literature Review

Much work has been done previously on different sign languages but not sufficient and effective work found in Urdu language. People find it difficult to communicate when they cannot speak or listen. We reviewed different research papers and inspected different technical methods to convert the different sign languages into text as shown in table 1. Due to the challenges faced, mostly work has been done on isolated images using CNN or on the languages whose datasets already exist by feature extraction whereas RNN used for video streams to align the sequence in end-to-end manner. Contrarily, Sign2Gloss2Text and media-pipe methods are also being used to convert the sign language into readable text.

3.2.1 CNN for Gesture Translation [1]

The paper is researched by the students of CS and Engineering department of Saranathan College of Engineering, India. They have used CNN model to make their system for gestures translation.

3.2.1.1 Summary of the research item

The methodology of proposed system is divided into four steps. First of all images are captured by using OpenCV python library. For each sign, approximately 3000 photos are gathered, which are then transformed to a csv file including the pixel values for more accuracy in predicting motions. Then n preprocessing step, the detected gestures are normalized by converting these frames to gray scale image and then the background is being eliminated by using algorithms.

The next step do classification of images. For this purpose, convolutional neural networks are used. CNN is a multilayer perceptron. Convolution Layer extract features from images by using image and kernel matrix and output of their dot product. Pooling Layer helps to normalize the size of convoluted image by using either max or min pooling. Flattening layer flatten the multi-dimensional matrix in order to be fed easily to the classifier. Next is Activation function. This system uses ReLU Activation function for achieving high accuracy. Final step is prediction which is done by using Convolution Neural Network.

3.2.1.2 Critical analysis of the research item

This system provides good precision with CNN to convert gestures to text but this accuracy decreases when the lightning is poor. That system captures gestures by only using hand moment but in explaining gestures, facial moments also play a vital role during communication. This system lacks in capturing facial and other body parts expressions.

3.2.1.3 Relationship to the proposed research work

In this system, signs are pre-processed after being webcam-captured. This system employs background subtraction at the pre-processing stage to remove the background. Therefore, any

dynamic background can be used by this system. Our project requires a dynamic background, and motions must be recognized over the entire upper body. This research report will be useful to us in achieving this goal.

3.2.2 Transformers for Sign Language Recognition [2]

The research was carried out by the students of University of Surrey, Germany. This study is done in the following pattern. Firstly, they have used the techniques CSLR and SLT in order to prove the effectiveness of mid-level glosses are effective. Secondly, they innovate the application of transformers. And finally they give their results to the future researches in that field.

3.2.2.1 Summary of the research item

The experiments in this study have carried out in the following ways. They have did two experiments. In their first experiment which is Sign2Gloss experiment, they uses Inception network uses CNN, LSTM and HMM models for learning SL after pre training. In the next experiment, they unified the recognition and translation task to examine the performance. For this purpose, they train their recognition model. After doing first experiment, they examine and shown that using pre-trained features, results outperform to the ImageNet bases spatial experiment. In second experiment, they have improved the SL recognition and relocation.

3.2.2.2 Critical analysis of the research item

After doing experiments they compare their results with the previous researches on that field. Their model surpass the performance of previous researcher's model for recognition and translation. They report a decrease of 2% WER by testing on both S2G and S2G2T setups. Furthermore, their setup of S2G2T surpasses previous text-to-text based Gloss2Text translation setup. Their experiment was not able to translate specific numbers and named entities like locations etc.

3.2.2.3 Relationship to the proposed research work

Mid-level glosses help to do effective translation. So this system will help us in achieving that efficiency.

3.2.3 Multi-layer CNN perceptron for Banks [3]

This article is researched by the students of CS and Engineering Department of Amrita Vishwa Vidyapeetham Amritapuri, India. They have designed a system to vanish the communication gap between the tongue-tied people and banks, by using CNN model features along with LSTM features.

3.2.3.1 Summary of the research item

The algorithm used by this system has four parts. Firstly, bank related dataset is created in the form of gestures sign videos using a mobile phone, every gesture was separately recorded multiple times. Then these videos are converted into image frames having a resolution of 1080x1920. Next step is of training the model. In this step, CNN is used feature extraction and LSTM is used to classify these gestures. CNN is multilayer perceptron. It has Convolutional layers. In this project, CNN is used only for feature extraction. Therefore, only a few layers of CNN are used for feature extraction. These layers are Maxpooling2D, conv2D and Average pooling. Last step is testing. In this phase, signs are converted to text.

3.2.3.2 Critical analysis of the research item

That system had a weaker side which was dataset collection. Due to the abate dataset for Indian Sign Language, the dataset was collected manually. The recorded videos were also of long length. This system provide 81% accuracy while dealing with everyday used words dataset and bank category dataset and that was the good side of system.

3.2.3.3 Relationship to the proposed research work

This article proposed a system which would be beneficial for the tongue-tied and hear-impaired individuals to do any of the bank procedures without using sensors or any other individual's help. This system is made by Indian citizens that is why they have used ISL for providing benefit to the Indian deaf and dumb community. As, there is no work done before on ISL. Similarly, there is no work done before on PSL, this way the system will help in our project.

3.2.4 Preprocessing Image using Kernel and its Classification [4]

The research on the article is conducted by the students of IS Department of Engineering College, Bangalore. They have discussed vision-based method to develop an android application. Videos are converted to frames then these images go for further processing and finally text is generated as an output of the system.

3.2.4.1 Summary of the research item

The algorithm is developed using Java-based OpenCV wrapper. The algorithm has 7 main steps. Calibration, frame processing, detection, kernel, dimensionality reduction, classification, and post-processing. In the first phase, skin tone is recognized. Only if the images pixel values in the certain bound of image classification, the frame will further go for classification else it will be discarded. Then they have used Gaussian blur for blurring image. Then the contours are found wherein skin color is present. In the next phase, an input image is read as BGR format then gives the image in RGB frame. Then these RGB images are down sampled and converted to grayscale. Then the Threshold contouring is applied to connect the separate paths and SVM algorithm is used to train the model.

3.2.4.2 Critical analysis of the research item

After comparing the combination of the above-mentioned array parameters, the average accuracy obtained was 98% and the confusion matrix will be made. Finally, this system becomes more efficient by using a customized SVM.

3.2.4.3 Relationship to the proposed research work

This paper has discussed the detailed image pre-processing techniques. They have discussed the 7 main steps of algorithm i.e. Calibration, frame processing, detection, kernel, dimensionality reduction, classification and post-processing. Our project needs to do pre-processing very efficiently. For this purpose, we will be consulting with this research paper.

3.2.5 Deep learning CRF techniques for CSLR [5]

This paper is a study of existing work on vision-based CSLR and deficiencies in existing work. It is researched by the students of Computer Science Department, and other schools of India.

3.2.5.1 Summary of the research item

This paper analyzes different techniques used in CSLR. Using CRF extracting features from a complex background a very crucial task because the results depend on these hand-engineered

features. Moreover, use of HMM and DTW are observed used on benchmark datasets. Hence, CRF handles the larger sequences in a better way than HMM standard corpora.

3.2.5.2 Critical analysis of the research item

This report is covering the challenges which can be faced in CSLR on benchmark databases and suggesting the better technique for sensor recognition.

3.2.5.3 Relationship to the proposed research work

The article has did comparison between different techniques used for gestures decoding or gestures translation. This study will help us in our project for selecting the good model which provides good accuracy and other functionalities.

3.2.6 SMTC-Transformer [6]

This paper is research by the member of Language Technologies Institute Carnegie Mellon University and LIX, Ecole Polytechnique Institute Polytechnique de Paris. This paper has introduced STMC-Transformer model for translating videos by filling the previous research gaps. And it has shown the better performance with using transformer instead of using gloss translation.

3.2.6.1 Summary of the research item

This system producing the results in the following strategies:

- **G2T** in which GT gloss annotations are translated for having perfect tokenization on both datasets used in that system. It is a text translation task. The original Transformer uses 6 layers for the encoder and decoder for NMT.
- **S2G2T** is a video-to-text translation which used STMC-transformer performed on PHOENIX-Weather 2014T. Firstly, they use the best performing model for German G2T to translate glosses predicted by a STMC trained network.

3.2.6.2 Critical analysis of the research item

The results of the experiment using the G2T and S2G2T models demonstrate that, although having low BLEU ratings, the translations are often of high quality. The fact that some outputs have slightly different word choices but the meaning of the phrase remains the same implies that BLEU is not a good representation of human qualities that are helpful for SLT. The dataset PHOENIX-Weather 2014T only contains weather forecast data, although more comprehensive data should be used to assess how well these models function. This was another flaw in that system.

3.2.6.3 Relationship to the proposed research work

This paper has revealed that glosses are a defective representation of sign language. This will help us in our research that how to use a better model and how to neglect the models which does not give good results. A key finding in that model is that by using a STMC network for tokenization better performance can be obtained, instead of translating GT glosses.

3.2.7 Recognition of Hand Gesture Movements with C3D Architecture [7]

This research project was done by the students at Saudi Arabia University, King Saud from different departments. Their project was funded by different Research Programs.

3.2.7.1 Summary of the research item

This paper has following contribution:

- C3D architecture optimizes the SL translation.
- Next step proposed an effective solution for the system, which mainly focuses on hand region.
- A framework named open pose is used for gesture segmentation.
- Next they optimizes the architectures for local features aggregation.

There are three primary steps in that system's algorithm. Processing of input data, feature learning, feature fusion, and feature classification. The input gesture videos are first converted to RGB. Next, the temporal dimension is normalized using linear sampling. In this step, manual cropping and normalizing are done further. This technique makes use of the open-source, deep learning-based open pose real-time human pose estimation framework for identifying each person's 2D key points in an image.

3.2.7.2 Critical analysis of the research item

The results of the proposed system is evaluated on different challenging dataset and telling how the system is giving effective results by performing on the proposed methods. But the proposed system was weaker in the area of implementation of Auto encoder. It was degraded when the depth of the architecture was increased.

3.2.7.3 Relationship to the proposed research work

Our project will be using deep learning algorithms. That research project has discussed CNN, C3D architecture in detail with their pros and cons, which help us in finding the correct solution for our project.

3.2.8 BERT Framework [8]

All ethical and experimental methods and procedures were approved for this research under Application No.1317 in 2018 by the University of Hong Kong Main Library Research Data Services.

3.2.8.1 Summary of the research item

This article is covering development of deep learning framework Sign BERT and multimodal vision of Sign BERT to translate Chinese sign language. For CSLR, ResNet is utilized to extract characteristics from video. A BLSTM layer is used as the second sequential module after the BERT model as the first sequential module. The videos from the corpus are slightly masked for pre-training the extraction of features module and the BERT model to increase the reliability of the Sign BERT framework.

3.2.8.2 Critical analysis of the research item

This model is more efficient and providing the better accuracy than the other state-of-the-arts achieving WER 23.30% which is much better than other models. It is using multiple layers of different models and techniques to convert CSL that makes it slower as there is too much processing in each layer. It is not providing any user friendly environment to help users who don’t know too much about IT.

3.2.8.3 Relationship to the proposed research work

Our research and this article are performing the same operation but on different datasets and sign language. It is using CSL and our model will be using PSL. This article used already collected corpuses to test and train their model. On the contrary we have to collect datasets to train and test our model as there is not enough data available for PSL.

3.2.9 GAN and 2D-CNN for CSLR [9]

This article was researched in different Visual Computing Lab at IT Institute of technologies.

3.2.9.1 Summary of the research item

In this article a model is designed to convert different sign languages to sentence. It is tested on Chinese, Greek and English sign language and their sentence generation. First, analyze the input frame sequence to derive spatial-temporal properties, 2D-CNN is used after which temporal convolution and pooling layers are used. Videos from datasets are resized to 256x256 and cropped to a fix size of 244x244, and up to 20% of videos frames were randomly removed because of memory restrictions. Videos containing more than 300frames were downscaled. BLST is used preserves the input information. Than a GAN based network classifies the outputs.

3.2.9.2 Critical analysis of the research item

It can translate multiple sign language. Chinese sign language, Greek and English. On the RWTH-Phoenix-Weather-2014 dataset, the CSL and GSL dataset, and the Signer Independent dataset, it is providing WER% of 23.4%, 2.1%, and 2.26%, respectively. This is much better state-of-the-art as it is performing on multiple languages. It does not provide any interface or application for normal users like a website or mobile application also it is not performing best on the English as there are models that are better than this in English Sign language.

3.2.9.3 Relationship to the proposed research work

This article is performing the same task we are hoping to achieve but in Pakistan sign language. The target or goal for this article as well as our work is same to dominate the communication gap between deaf and normal people.

3.2.10 Deep Learning models LSTM and GRU [10]

This article was researched under different Departments of Electronics, Engineering, Communication, and technology in Osaka Japan.

3.2.10.1 Summary of the research item

This article describes a technique for translating Indian sign language to words. It uses deep learning models to detect and translate gestures, LSTM and GRU which are feedback based leaning models. Feature vectors are extracted using InceptionResNetV2 model is used to extract features from videos and then they are further processed by passing them from different layers. There are total 4 different layers of GRU and LSTM two for each one. 97% accuracy is achieved by single layer of LSTM followed by GRU. It uses dataset IISL2020 and personal collection of data.

3.2.10.2 Critical analysis of the research item

The best thing that this article is covering is a model to translate Indian sign language. One of the main contribution of this article is data collection as there is very little data for Indian sign

language but there are other articles and models that perform well than this one they even translate to complete sentence and give much better accuracy.

3.2.10.3 Relationship to the proposed research work

Just like this article we are designing the same system for our Pakistan sign language and we will be facing the same difficulties like very small dataset for training and testing purpose. It is using deep learning techniques that will help achieve good results on small dataset.

3.2.11 Multimodal Sensor Fusion [11]

This article was received on 20 October 2020 and was accepted on 29 December 2020. Publishing date of this article is 22 January 2021.

3.2.11.1 Summary of the research item

The models and systems designed to recognize gestures and translate them face a main challenge is to recognize sign in continuous SL videos. This article mainly tries to solve this by introducing computer vision-based system. The technique uses the signer's skin tone to identify their hands and head, therefore they must wear long sleeves. Features are extracted using different techniques. HMM is used to extract, recognize, and classify signs. It reached 2.24% and 2.9% advancement on one and two hand motions, obtaining 95.18% recognition accuracy for a single gesture and 93.87% for two. As a dataset, it makes use of gathered sign language motions. 33 different signs were captured simultaneously utilizing the Kinect sensor and Leap motion, 22 of which were signed with both hands, and the remaining signs with just one.

3.2.11.2 Critical analysis of the research item

The article is mainly focused on detecting word end boundaries and hand and head detection in sign and it achieved a good accuracy. On the other hand, it has limitations like skin color and wearing a long sleeve shirt is compulsory. Skin color can affect the accuracy and data set is too small.

3.2.11.3 Relationship to the proposed research work

Detection of hand gestures as well as head and maybe other parts of body is most important thing in sign language translation system. This model will be a part in recognizing signs.

3.2.12 Using Multiple Trademark Points [12]

This article was added to IEEE Xplore in June 2022.

3.2.12.1 Summary of the research item

Most of the continuous SL recognition techniques use images as input that can give partial results and facial expressions are also not considered in it. This article solve the problem and characteristics of joints and bones are extracted using a residual spatial temporal adaptive graph convolutional network. In this article GRU encoder is used to capture short and long term dependence, and decoder and attention mechanism are used to calculate encoder information. CSL dataset and PHOENIX-2014 dataset are used to verify this model. The word error rate on CSL dataset is 2.2 and 29.4, and on PHOENIX-2014 dataset it is 23.4 and 23.8.

3.2.12.2 Critical analysis of the research item

This model is performing well in term of word error rate. It is including joints and bones and also the facial expressions. On the other hand, this model is much slower than previous researches and other models like BERT based, because of that this model is slower in real time translation and difficult to entertain real time translation.

3.2.12.3 Relationship to the proposed research work

The extraction of face features and position of hands relative to other parts of body will be helping our system to detect signs more accurately and achieve better results.

3.2.13 Multi-Scale Local-Temporal similarity Merging [13]

The article was researched in different universities of China such as AI Lab of Xiaomi, Software institute Sciences academy.

3.2.13.1 Summary of the research item

In most of the models to Generating sentence or recognizing sign language order of gloss sequence is essential. Information about the fine-grained gloss level must be recorded. This study addresses this issue through temporally similarity-based adaptive fusion of local features. It uses Multi-scale Local-Temporal Similarity Fusion Network. Video frame sequences are first encoded using the frame feature extractor and outputs series of spatial features. 1D-CNNs are used by gloss feature encoder across the temporal dimension after extracting the spatial features to learn context. Such an encoder has 2 levels, each of which is layered with several 1D-CNN layers and performs a separate purpose. Then this software is used to learn the local temporal similarity and it has three main components Multi-scale, Feature analyzer and Position-aware Temporal Convolve to overcome drawbacks of CNNs and improve the temporal consistency of each combined representation. It uses Connectionist Temporal Classification (CTC) for sequence learning. The model uses RWTH-PHOENIX-Weather 2014 datasets (RWTH) and achieved the word error rate (WER %) 23.8 on development and 23.5 on testing.

3.2.13.2 Critical analysis of the research item

This model identified drawback of fully convolution network for sign language recognition and that CNNs become agnostic to similarity and temporal consistency. It performs quite well but it loses the word error rate as other models has lower word error rate.

3.2.13.3 Relationship to the proposed research work

It's crucial in the majority of models to produce sentences or recognize sign language in order. So this model might help us in future for our work in extracting glosses without losing essential information.

3.2.14 Two-stream transformer [14]

This article is braced by the China Research Center of Engineering Technology and National Natural Science.

3.2.14.1 Summary of the research item

This article is covering the method of two stream light weight transformer to translate the Chinese continuous sign language of hands and mouth movements into text by obtaining the light RGB frames and extracts the spatial information with static and dynamic features of the body with the Spatiotemporal context feature. This project is based on two segments; video

extraction with 2D CNN and sign language translation by training the model in end-to-end fashion. It is using the Chinese sign language corpus of 3080 high quality videos generated by professionals.

3.2.14.2 Critical analysis of the research item

This model is more efficient and providing the better accuracy than the previous models but the dataset generated is not large enough to cover all of the domains. Moreover, it includes multiple steps to translate sequence-to-sequence with multiple encoder and decoder layers which makes it slower. Furthermore, there is no mobile terminal view and extensive services of this model.

3.2.14.3 Relationship to the proposed research work

Unlike today, there was no such Chinese sign language corpus which is generated with the help of mute-deaf people in schools and colleges and embedded in the software to train system. Similarly, we ourselves have to generate our corpus of Pakistani Sign language for our system. This research is directing us the way to extract features from videos with RGB frames and get maximum accuracy.

3.2.15 Multi-view data using GCN and Transformers [15]

This article based on Multi-view Learning Guest which belongs to the Collection of Special Issue. The editors are Mr. Chao et al.

3.2.15.1 Summary of the research item

Traditional Sign language recognition used CNN and RNN to extract the spatial and secular features respectively. This research depends on the major three steps: spatiotemporal with multi-view immersing network, translation of sign language and proposed the GCN and transformers. Transformer encoder network layer takes the sign language of RGB clip and SKE (skeleton) frame and used the fully connected polling layer and sends it to the transformer decoder network after CTC which finally produces the spoken text.

3.2.15.2 Critical analysis of the research item

It provides the high precision accuracy but failed to learn the long-term dependencies. It does not introduce bottlenecks information but the parallel training data in transformers increases the computational efficiency.

3.2.15.3 Relationship to the proposed research work

This embedding Network helps to learn the RGB and skeleton data frames directly in spatiotemporal. Thus, if we extract the features in skeleton and RGB frames then this neural network will help us in feature extraction as this network is used to enhance the vigor and lessen the detrimental impact of unspecified joints recognition.

3.2.16 3D transition frame image with Gesture Tracking [16]

This work was reinforced in 2021 by Dongguan City College with the help of Young Teacher Development.

3.2.16.1 Summary of the research item

In this research, the hand shape and movements of hands assisted with body parts is being recognized and translated with the combination of optical flow algorithms and particle filters.

A 3D smooth transition frame image is obtained with hardware equipment and virtual reality modeling language (VRML).

3.2.16.2 Critical analysis of the research item

It concretely analyzes each step and extract the features but failed to produce the phenomena of ‘particle degradation’ in which the individual calculation of the particle becomes zero when iterations increase. Moreover, it fails to provide the facial expressions with the hand movements which play a crucial role in producing effective sign language.

3.2.16.3 Relationship to the proposed research work

This method can be used to capture the 3D image frames and applies filters on it to get the maximum proficiency. Secondly, it covers the upper half of the body to track the gestures of hands interacting with the body which we will use for our system to generate the text of the human anatomy such as abdomen, forelimb etc.

3.2.17 CNN and RNN for video segmentation [17]

The support from the Deaf association of Ouarzazate is given to generate the dataset of Arabic Sign language.

3.2.17.1 Summary of the research item

Arabic sign language video is classified in image sequences with the help of two different models of deep learning. One is 2DCRNN (two dimensional convolution neural network) and second is 3DCNN (three dimensional convolution neural network) which are being used for spatiotemporal feature extraction and tested with fourfold cross validation with minimum dataset to train the system.

3.2.17.2 Critical analysis of the research item

The accuracy achieved in feature extraction is very high but this model cannot be trained with the anomaly data as well as it does not provide any sentence level sequence.

3.2.17.3 Relationship to the proposed research work

As discussed earlier, we have limited dataset available in Pakistani sign language and this model is helpful as deep learning techniques provided gives a good classification results with minimum training samples.

3.2.18 Baseline Multi-Modality [18]

The research article is written with the contribution of many authors which covers the simple way of translating sign language.

3.2.18.1 Summary of the research item

Human actions are the basic part to be identified in gestures recognition. This model is useful for pre-training the system with gloss-to-text within the corpus's domain and sign-to-gloss on the overall influence. For joining these two networks, a visual language mapper is used via neural machine translation and creates the baseline for further researches to covert the sign language into text.

3.2.18.2 Critical analysis of the research item

An elementary solution is provided for sign language translation but does not give maximum cross entropy accuracy in sequence-to-sequence learning model. Besides, it does not train itself on the external dataset.

3.2.18.3 Relationship to the proposed research work

This article shows a simple yet effective and directed way to convert the sign language into text. It clearly shows the efficient way to train our model with the dataset and helpful in the isolated videos similar to our dataset .In addition to this, it pre train the visual and language models from general to within domain such as we are coving the medical domain.

3.2.19 Media-Pipe GRU model [19]

This project was funded by National Research Foundation of Korea which was received in February 28, 2022 and accepted by the start of July 2022.

3.2.19.1 Summary of the research item

Precisely identifying the gestures before translating them into text is the critical section which is proposed in this article. This model is using three techniques which include data preprocessing of feature extraction with Media-pipe framework. After removing the null points, the data saved in file with labeling and built-in data augmentation is used to landmark the body parts and finally the gestures are being recognized by training the data in Media-pipe optimized gated recurrent unit (MOPGRU) model which generates the text.

3.2.19.2 Critical analysis of the research item

A limited dataset of the sign language videos is being used but with the faster coverage speed and maximum precision accuracy. Moreover, it does not work for continuous sign language.

3.2.19.3 Relationship to the proposed research work

The important thing in feature extraction is to remove the unwanted material from the video. This model plays a vital role in giving an effective way to extract the relevant information by detecting landmark on the human body. Identifying Landmarks is the important feature in knowing about the exact movement of gestures which we have to propose in our research.

3.2.20 Iterative training using Deep Learning [20]

The students of Tsinghua University and Peking University of China have contributed to the proposed project which translates the real-time videos into text. Their research interest was computer vision and deep learning.

3.2.20.1 Summary of the research item

The proposed model is mainly covering the end-to-end long continuous sign language video stream and converting them into video segments in the form of RGB frames and visualization of optical flow. It then extracts the features by adopting the deep learning technique CNN with time-stacked fusion and aligns the sequence by assigning them gestural labels with the help of Bi-LSTM. This process repeats iteratively and improves the architecture.

3.2.20.2 Critical analysis of the research item

Sequence-to-sequence translation is a challenging task which has been catered in this iterative optimization process with the limited training dataset. In contrary to this, simultaneous related

channels need more attention in gestures recognition which has not been covered in this method.

3.2.20.3 Relationship to the proposed research work

Deep Learning techniques are widely being used in translating the gestures into text or articulation. In this project, every step of the module is being discussed in detail which will oblige us in our research from applying the neural network in temporal feature extraction to the gestural labels and generating text in sequence.

3.3 Literature Review Summary Table

Following is the table providing findings of past articles related to gestures translation.

Table 1: History of Gestures Decoding methods

The summary of various methods of converting gestures to text or speech in the past from 2020-2022 is presented here.

No.	Name, reference	Author	Year	Input	Output	Description	Accuracy/ Result
1.	Sign Language Translation, [1]	Harini R et al.	2020	Images of signs	Spoken language sentences in the form of text.	The aim of this system is to eliminate communication barrier between common people and deaf people without need of any specific color background, hand gloves or any sensors.	Accuracy: 99.91%
2.	Sign Language Transformers: Joint End-to-end Sign Language Recognition and Translation, [2]	Necati Cihan et al.	2020	Sign language videos	Spoken language sentences in the form of text.	They have given SL Transformers. It is a transformer which jointly learn recognition and translation in sequential manner.	Result: They decreased a WER by 2% in testing on both S2G and G2T setups
3.	Mudra: Convolutional Neural Network based Indian Sign Language Translator for Banks, [3]	Gautham Jayadeep et al.	2020	Indian Sign Language videos	Word by word translation into spoken language text.	The objective of this system is to propose a method to convert signs patterns in Indian Sign Language dictionary related to the bank into text.	Accuracy: 81%
4.	Real-time Conversion of Sign Language to	Kohsheen Tiku et al.	2020	American Sign Language sign	Translated text	The objective of this paper is to develop an android application for converting signs to text/speech without	Accuracy: 98%

	Text and Speech, [4]					using sensors and completely free of cost.	
5.	Understanding vision-based continuous sign language recognition, [5]	Neena Aloysius and M. Geetha	2020	Continuous sign sequences	Spoken language text or speech.	This paper is a study of existing work on vision based CSLR. It uses the sensor-based systems and benchmark databases.	Accuracy: 80-90%.
6.	Better Sign Language Translation with STMC-Transformer, [6]	Kayo Yin and Jesse Read	2020	Gloss2Text (G2T): Text Sign2Gloss2Text (S2G2T): Video	Gloss2Text (G2T): Text Sign2Gloss2Text (S2G2T): text	This paper is performing the translation of video into text on mid-level glosses.	Result: STMC obtains a better WER of 21.0
7.	Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation, [7]	Muneer Al-Hammadi et al.	2020	Sign	Text	This method is used to efficiently eliminate the extra background.	Accuracy: 87.69%
8.	Sign BERT: A BERT-Based Deep Learning Framework for Continuous Sign Language Recognition [8]	Vinvent W.L.Tam et al.	2021	Videos of Sign language	Text	Two methods are being used simple Sign BERT and multimodal vision of Sign BERT. To extract characteristics from videos for continuous SL recognition (CSLR), Residual neural network (ResNet) is deployed. Multimodal Sign BERT is giving much better performance.	Result: 23.30 WER% result on unseen sentence test of CSL
9.	Continuous Sign Language Recognition through a Context-Aware Generative Adversarial Network, [9]	Kosmas Dimitropoulos et al.	2021	Videos of Chinese ,German and English Sign language	Text Sentences	Prior to using temporal convolution and pooling layers, 2D-CNN is utilized to extract spatiotemporal properties from the input frame sequence. Videos from datasets are chopped to a fixed size of 244x244 and scaled to 256x256. Up to 20% of the video frames may also have been arbitrarily eliminated due to resource	WER of 2.1%, 2.26% and 23.4%

						constraints. Videos with more than 300 frames were shrunk. The input data is preserved when BLST is employed. The output is then classified using a network powered by a Generative Adversarial Network (GAN).	
10.	Deepsign Sign Language Detection and Recognition Using Deep Learning, [10]	Chintan Bhatt et al.	2021	Video sign in Indian sign Language	Words	Video features are extracted as feature vectors using the InceptionResNetV2 model, which are then processed further by being passed through several layers and finally uses GRU and LSTM for translating text.	Accuracy: 97%
11.	Vision-based continuous sign language recognition using multimodal sensor fusion, [11]	Mohamm ed Jemni et al.	2021	Video containing signs	Sentences	The technique uses skin tone to identify the signer's hands and head, thus they must wear long sleeves. Using Hu moments Hu, Zernike moments Zernike and Fourier descriptors Zahn and Roskies techniques, features are retrieved. To extract, identify, and categories signals, Hidden Markov Model (HMM) is utilized.	Accuracy: 95.18%
12.	Continuous Sign Language Recognition Using Multiple Feature Points [12]	Yanliang Jin et al.	2022	Video containing signs in Chinese and English	Sentence	To extract characteristics of joints and bones, the author of this paper designed a RST-AGCN, which can identify the location of hands in relation to other body components. The GRU encoder is employed in this article to capture both short and long terms dependencies, and the decoder and attention mechanism are utilized to calculate encoder information	Accuracy: 91%
13.	Multi-Scale Local-Temporal Similarity Fusion for Continuous Sign Language Recognition	Pan Xie et al.	2021	Video of German Sign Language	Sentence	A first frame feature extractor from segments encodes after extracting the spatial features, the gloss feature encoder employs 1D-CNNs across the temporal dimension to learn context	They achieved the word error rate (WER %) 23.8 on developme nt and 23.5 on testing.

14.	Two-stream lightweight sign language transformer,[14]	Yuming Chen et al.	2022	Videos of Chinese Sign language and convert them into RGB frames	Chinese Translated text	Two segments are being used in this model which includes the feature extraction and training data set and generating sequence to sequence sentence. It also features the model on real-time videos.	Accuracy: 96%
15.	Sign language recognition and translation network based on multi-view data, [15]	Ronghui Li et al.	2022	Multi-view input sign data	Spoken text sentences	This model is tested on different datasets of Chinese and German sign language and gives the high accuracy than the gloss architectures.	Results: 22.80 BLUE-4 scores with CSL-daily
16.	Research on the improved gesture tracking algorithm in sign language synthesis. [16]	Quanyu Song and Rong Lu	2022	Sign Language video	3D improved sign language realistic frames of videos	It is used to improve the sign language video segments and covering the upper part of the body.	Result: It gives the accurate results maximum in the generation of fourth frame.
17.	Isolated Video-Based Arabic Sign Language Recognition Using Convolutional and Recursive Neural Networks, [17]	Mustapha Kardouchi et al.	2021	Arabic Sign Language isolated videos	Image features classification	Image characteristics in Arabic are taken from the collection of isolated videos using fully connected network and trained with four cross validation on the dataset of 244 videos consist of 56 signs.	Accuracy: 98%
18.	A Simple Multi-Modality Transfer Learning Baseline for Sign Language Translation, [18]	Zhirong Wu et al.	2022	Sign language Video	Text	A useful model is described for sign language translation which can further be used for different datasets by training the model with Sign2Gloss and Gloss2Text.	Results: They achieved 26.95 BLEU-4

19.	An integrated media pipe-optimized GRU model for Indian sign language recognition, [19]	Subramanian Barathi et al.	2022	Indian Sign language video	Translated Hindi text	Mediapipe holistic is used to landmark the whole body parts shown in the web camera and based on the trained dataset, it identifies the gestures and convert it into text	Accuracy: 95%
20.	A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training, [20]	Changshui Zhang et al.	2019	Continuous Sign Language Recognition videos	Sequence to sequence sentence in text	Mediapipe holistic is used to landmark the whole body parts shown in the web camera and based on the trained dataset, it identifies the gestures and convert it into text.	Result: The value of WER reduced 2.80%

3.4 Conclusion

In this study, we offered many associated research models for sign language translation and recognition to facilitate the special people. It includes several steps, to extract the spatiotemporal features extraction, traditional sign language recognition used CNN and RNN [17] [14] with the addition of transformers. Besides, VRML is used to get the 3D smooth transition [16] frame images and Sign2Gloss2Text is used to pre train the system within the domain [18]. In addition to this, an efficient way MOPGRU is used to translate the gestures by identifying landmarks [19].

The suggested methods are for different languages but generating text in Urdu language and its segmentation is critical task which will be catered in this research. We can use the existing models after preprocessing the data and produce the Urdu text by normalizing and training the dataset. It is directed to further work in future for generating end-to-end long sentences in Urdu.

3.5 Selected Model

After studying all the model we came to a conclusion that LSTM is the best model for our scenario. LSTM (Long Short-Term Memory) networks and dense layers can be used effectively for sign language translation tasks. LSTM networks are a type of recurrent neural network (RNN) that can model sequential data, making them well-suited for tasks involving time-dependent patterns such as sign language. When combined with dense layers, which are fully connected layers, LSTM networks can learn complex relationships and capture important features from the input data. The dense layers can help map the learned features to the output classes or translations in the sign language translation task. MediaPipe Holistic is a computer vision solution that provides various pre-trained models for human pose estimation, hand tracking, and face detection. It can be used to extract visual features from sign language videos or images, which can then be fed into the LSTM and dense layers for translation.

By combining LSTM networks with dense layers and MediaPipe Holistic, we are creating a system that takes sign language videos or images as input, processes them, and translates them into Urdu text.

Chapter 4: Software Requirement Specifications

The SRS document is a guide which provides the complete description for the project. In this document, we have specified the functional, nonfunctional, hardware and software requirements. Moreover, use cases with detailed description, graphical user interface and database are being proposed in this document.

4.1 List of Features

- A user can start capturing the gestures after starting the system
- User can switch between pausing or resuming for capturing the features

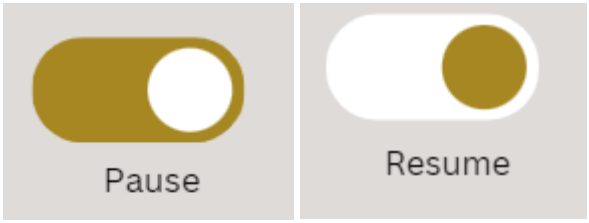

Figure 2: Pause/Resume Button

This figure represents the buttons for Pausing and resuming the video

- A user can use the already existing videos of the words in the database or capture them in real time
- An interactive graphical interface for the user to understand the system easily
- Buttons are mentioned in the navigation bar to go into the perspective page
- Our system will generate the text in Urdu after selecting the button generate text

4.2 Functional Requirements

All the external features of the system are being discussed in this section in which the user can start, pause, resume and discard capturing the videos.

1. **Sign Up:** The system shall allow the new user to create an account by filling out the information required.
2. **Log In:** The system shall allow users to login their account by providing credentials and use the software.
3. **Web Camera access:** The system shall allow the user to access the camera to capture the signs.
4. **Start Capturing:** The user shall be able to start capturing the signs.
5. **Uploading Video:** The system shall allow the user to upload the selected video from the database.
6. **Cancel/Discard Sign:** If any sign is captured wrongly, the user shall be able to cancel it and record it again.
7. **Pause Video Capturing:** The user shall be able to pause the recording video.
8. **Stop Video Capturing:** The user shall be able to stop the recording video.
9. **Resume Video Capturing:** The user shall be able to resume the paused video.

10. **Text Generation:** The user shall be able to generate the text. Then, text will be generated in Urdu according to the signs.

4.3 Quality Attributes

The quality attributes of the system includes its performance, usability, performance and availability. The system will provide security to the user such that their real time videos will not be used anywhere else. It will be available 24/7 and an easily understandable graphical user interface is being provided. These attributes will help to enhance the functionality and engage the multiple users to benefit from the services of our system.

4.4 Non-Functional Requirements

This section contains the non-functional requirements to measure the quality of the system which includes the terms such as time of availability, reliability, usability, serviceability, performance and security of the system.

- **Availability:** The services of the system will be available 24/7
- **Reliability:** The system is reliable and fast as it is connected with a local host and does not crash while producing the video
- **Usability:**
    - Multiple users shall be able to use the system at a time
    - The outputs shall clearly specify the result of the query in simple words
    - Any failure to basic operations shall result in an appropriate error message
- **Serviceability:**
    - The system shall provide the real time capturing services of the videos and generating text
    - The system shall allow the user to access the videos saved in the database and use them
- **Security requirements:**
    - The system shall allow the users to have a unique email/ phone number and passwords of their accounts from which the only particular user can access his account.
    - The system shall not be able to save the real time captured videos of the user
- **Performance requirements:**
    - System shall be able to serve up to 1000 users at a time
    - Up to 95% of the transactions made to the DB shall result in less than 3 seconds which includes:
        - Response time from the data center shall be less than 2 seconds
        - Error rate of the system shall be less than 6%

4.5 Assumptions

The assumptions made for the specification for the user to use the system are listed as:

- Users shall know the Urdu Sign language to generate the signs

- Users shall know the basic English to use the system as the language of our system is English
- The user shall have the basic knowledge of technology to use the system
- The user shall be comfortable in capturing the sign language video.

4.6 Hardware and Software Requirements

To develop and deploy the project, hardware and software requirements needed are being discussed in this section.

4.6.1 Hardware Requirements

The hardware requirements to successfully use the system at the user end needs:

- **Any portable equipment:**

  A portable device is needed to carry it to the doctor's place.

- **Memory in the system:**

  The device shall have the sufficient space available to setup the system.

- **Camera available:**

  The user shall have the camera in the device to capture the signs.

- **Internet Connection:**

  Internet connection is required to set up the system and the user can use the system offline afterwards.

4.6.2 Software Requirements

The software requirements for the developers and the users are as:

- **Developers Software Requirements:**
  - Jupyter Notebook/ Visual studio Code
  - Python libraries
  - Microsoft SQL Server
  - Internet
  - Draw.io Software
- **User Software Requirements:**
  - A software to run the system

4.7 Use Cases

This section lists use cases or scenarios from the use-case model if they represent some significant, central functionality of the final system, or if they have a large architectural coverage—they exercise many architectural elements or if they stress or illustrate a specific, delicate point of the architecture.

4.7.1 Login

Name	Login		
Actors	Patient		
Summary	The user must fill out the login form using their email and password in order to access the system, which will then take them to their account home screen.		
Pre-Conditions	For logging into the system, the database entries must already include that user either manually inserted by a developer or added by any authorized user. The user must not already be logged in.		
Post-Conditions	The user's session has been successfully formed, and they will be sent to the home page of their account.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	The login page is opened by the user.	2	The login screen, which requests an email address and password, is shown.
3	User input valid password and email.	4	The system authenticates the user's email address and password, creates a user session, and delivers them to the home page.
5	The user clicks the "forgot password" section after forgetting their password.	6	The user is led to another form where he must input his email address.
7	The user fills up the email address linked to his account.	8	The system sends the OTP after verifying the email by checking the database.
9	User inputs a valid OTP.	10	Users are directed to the input areas to enter a new password when the system has verified their OTP.
11	The user inputs his new credentials on the login page after being sent there.	12	The system confirms and takes the user to the main page of his account.
Alternative Flow			
3	User inputs an incorrect email address or password.	4-A	*Incorrect email address or password entered,* the system answers with an error message.
7	The user types in the wrong email address.	8-A	The system replies with an error. “*This email is invalid.”*
9	The user types in the wrong OTP.	10-A	*"This OTP is invalid,"* the system answers with an error message.

4.7.2 Signup

Name	Signup
Actors	New User/ Patient

Summary	The user must provide his name, email address, password and phone number in the input boxes before a user account may be established if the information is all accurate.		
Pre-Conditions	The user must fill out the whole sign-up form and ensure that none of his information—including username or email—exists in the database.		
Post-Conditions	After successfully creating the user's account, the user is sent to the login page.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	The user opens the sign-up page.	2	The signup page is presented as a form with all required input fields (name, email, password, phone number) visible.
3	The user fills up all required fields with accurate data.	4	The system checks the information entered into the forms before directing the user to the login page.
Alternative Flow			
3	The user enters incorrect input fields.	4-A	*"Invalid data on the associated fields,"* the system's response reads.
3	Any necessary input fields are omitted by the user.	4-B	When an error occurs, the system displays the message *"Please fill in the missing field(s)."*

4.7.3 Upload Video

Name	Upload Video		
Actors	Patient		
Summary	The user clicks on the upload video button to upload the video from the database and the video will be loaded into the system.		
Pre-Conditions	The user must be logged in into the system.		
Post-Conditions	After uploading the video, the user will be given a text generating option.		
Special Requirements	The user must have provided memory access to the system.		
Basic Flow			
Actor Action		**System Response**	
1	The user clicks the upload video button.	2	System asks for memory access.
3	The user will provide memory access.	4	Video will be loaded into the system from the user’s memory.
Alternative Flow			
3	The user will not provide memory access.	4-A	Video will not be loaded into the system and it will generate an error

			message *"Cannot access to the memory"*

4.7.4 Real Time Capture

Name	Real Time Capture		
Actors	Patient		
Summary	The user clicks on the Real Time Capture button to capture the video in real time. After pressing the button, the camera will start capturing the video and the video will be loaded to the system.		
Pre-Conditions	The user must be logged in into the system.		
Post-Conditions	After uploading the video, the user will be given a text generating option.		
Special Requirements	User must allow camera access.		
Basic Flow			
Actor Action		**System Response**	
1	The user clicks the Real time capture button.	2	System will ask for the camera access.
3	The user will provide camera access to the system.	4	System will start capturing video from the camera.
Alternative Flow			
3	The user will not provide camera access to the system.	4-A	System will not be able to capture the video and it will generate an error message *"Cannot access to the camera"*

4.7.5 Cancel Button

Name	Cancel Button		
Actors	Patient		
Summary	The user will press the cancel button for discarding the captured video of signs.		
Pre-Conditions	The user must be logged in and have some captured video of signs.		
Post-Conditions	The user will be given an option to capture the video again.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	The user will press the cancel button.	2	System will discard that captured video.
Alternative Flow			
None			

4.7.6 Pause Button

Name	Pause Button		
Actors	Patient		
Summary	The user will be able to pause the real time video being captured.		
Pre-Conditions	The user must be logged in and must be capturing some real time video.		
Post-Conditions	The user will be given an option to resume the paused video.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	The user will press the pause button.	2	System will pause the video and will allow the user to resume capturing again.
Alternative Flow			
None			

4.7.7 Resume Button

Name	Resume Button		
Actors	Patient		
Summary	The user will be able to resume the paused real time video being captured.		
Pre-Conditions	The user must be logged in and must have some paused real time video.		
Post-Conditions	The user will be given an option to stop capturing video.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	The user will press the resume button.	2	System will resume the video.
Alternative Flow			
None			

4.7.8 Stop Button

Name	Stop Button
Actors	Patient
Summary	The user will be able to stop the video being captured.

Pre-Conditions	The user must be capturing some video.		
Post-Conditions	The user will be given a text generating option.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	The user will press the stop button.	2	System will stop the video.
Alternative Flow			
None			

4.7.9 Generate Text

Name	Generate Text Button		
Actors	Patient		
Summary	The user will be able to generate the text after capturing the video by using any of the forms.		
Pre-Conditions	The user must have some captured video.		
Post-Conditions	The user will be given a text as an output.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	The user will press the Generate Text button.	2	System will generate the text.
Alternative Flow			
None			

4.8 Graphical User Interface

The only user of our system is the hearing impair person who has to get his checkup done from the doctor and needs to translate his gestures into text. The graphical user interface is shown for different pages with the description of their work. In the header, the navigation bar contains the information of the system which includes home, capture, upload video

4.8.1 Home Page

The home page will be open at the user end after getting into the system.

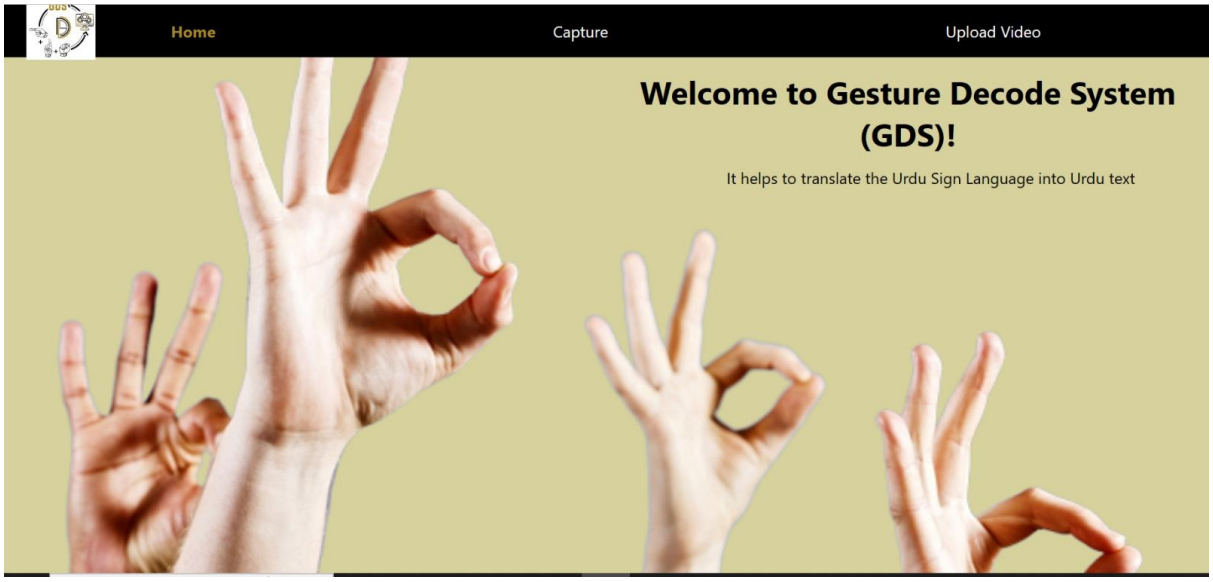

Figure 3: Home Page
This figure represents the home/main page of our system

4.8.2 Signup Page

A new user can register by creating an account and can become a permanent user of the system by providing the username, email address, phone number and a valid password.

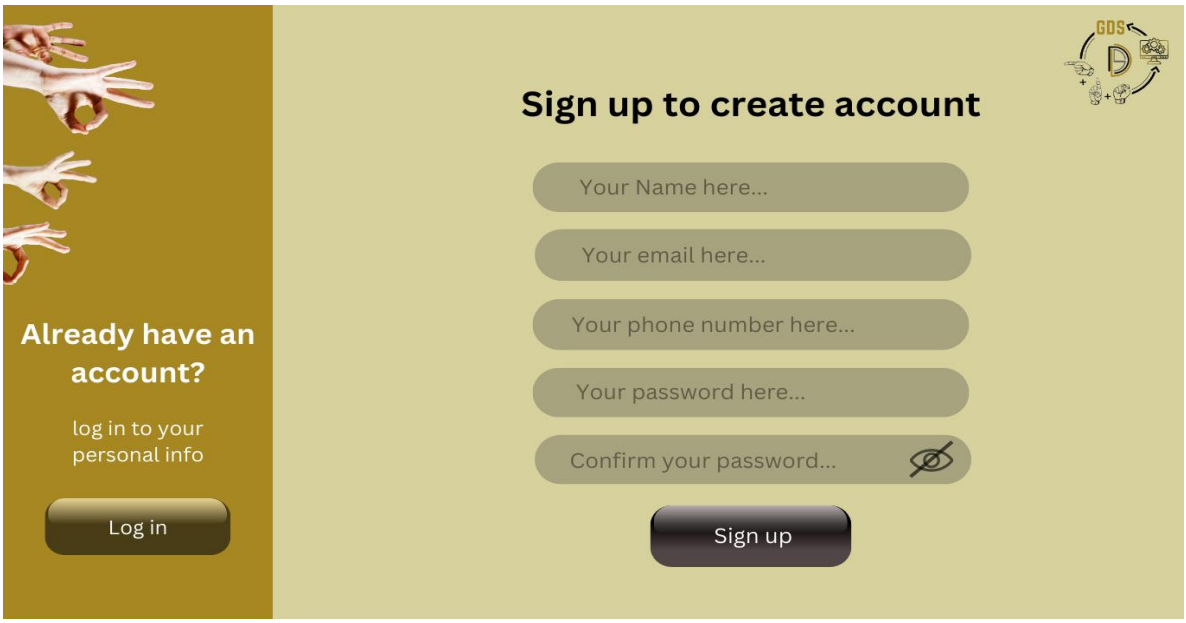

Figure 4: Signup Page
This figure represents the signup page of our system

4.8.3 Login Page

The user can login into the system if an account is created already. A unique ID is being assigned to every user and a record is maintained.

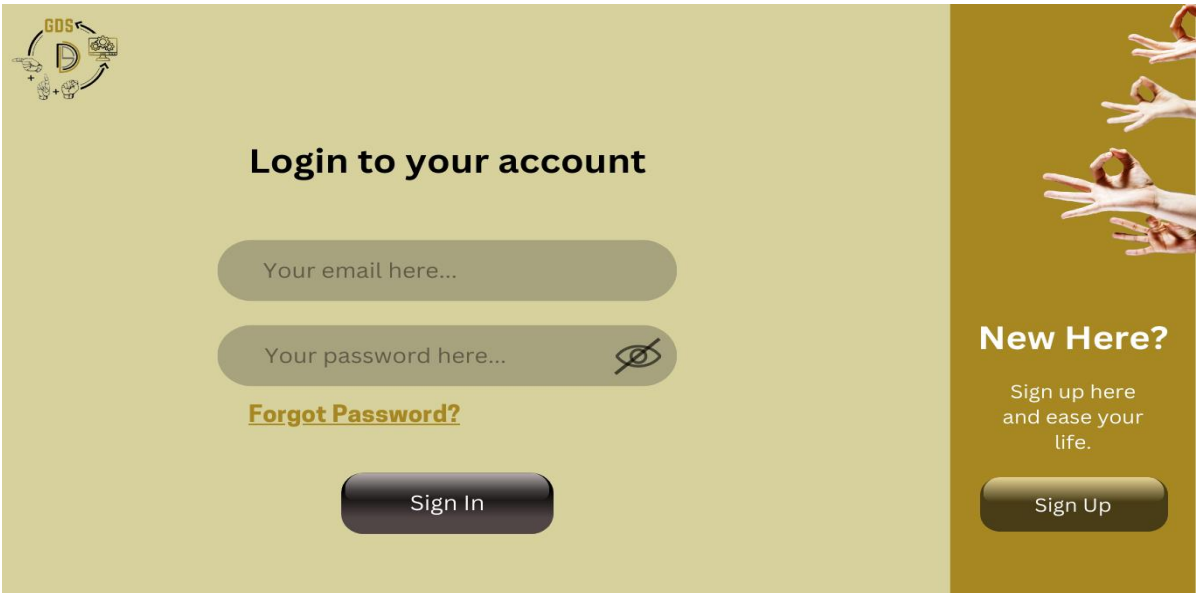

Figure 5: Login Page
This figure represents the login page of our system

4.8.4 Capture Videos

The following page will appear after selecting the capture option from the main page. In this page, a user can record the real time video and generate text. A user can discard any wrong sign, pause and resume the video, stop the recording and can generate text after completion of the video.

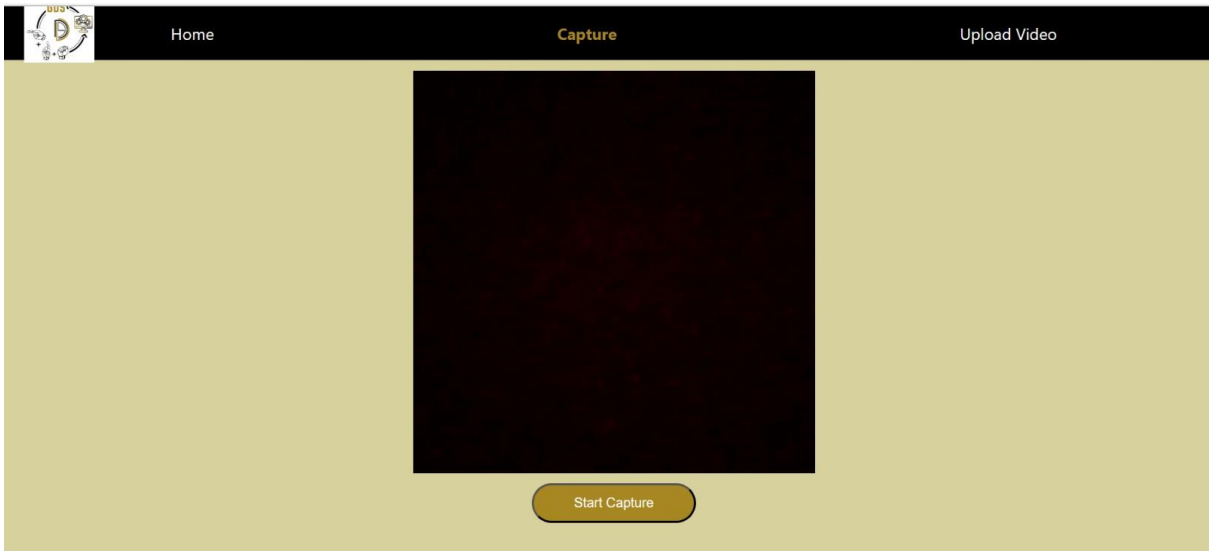

Figure 6: Capture Videos
This figure represents the capturing videos page of our system

4.8.5 Upload Video

The following page will appear after selecting the upload video option from the main page. In this page, all the videos of the words available in the database of the system will be shown to the user and he can then select or extract the specific video and generate text.

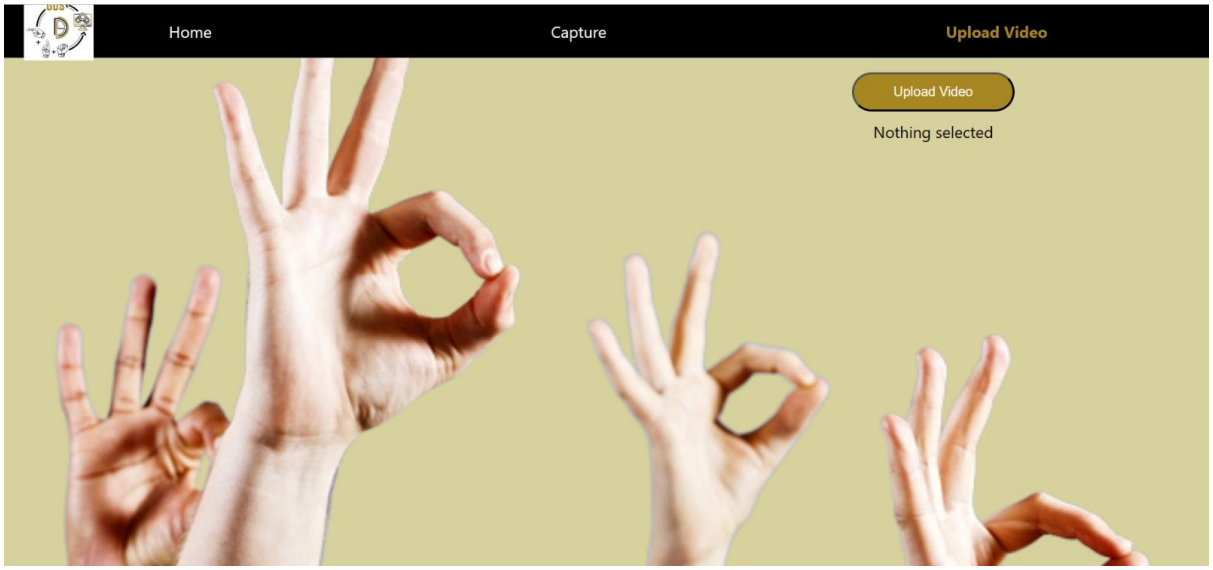

Figure 7: Upload Video
This figure represents the uploading videos page of our system

4.9 Database Design

We designed our system with the help of ER Diagram and provided the data information of users is provided in Data Dictionary

4.9.1 ER Diagram

The relation of entities User, login, sign up, upload video and real time capture are shown with their relationship in crow foot notation such as a user can have one and only one account, a user can select zero or many videos to upload the existing videos or can capture the real time videos and generate text as shown in the figure below.

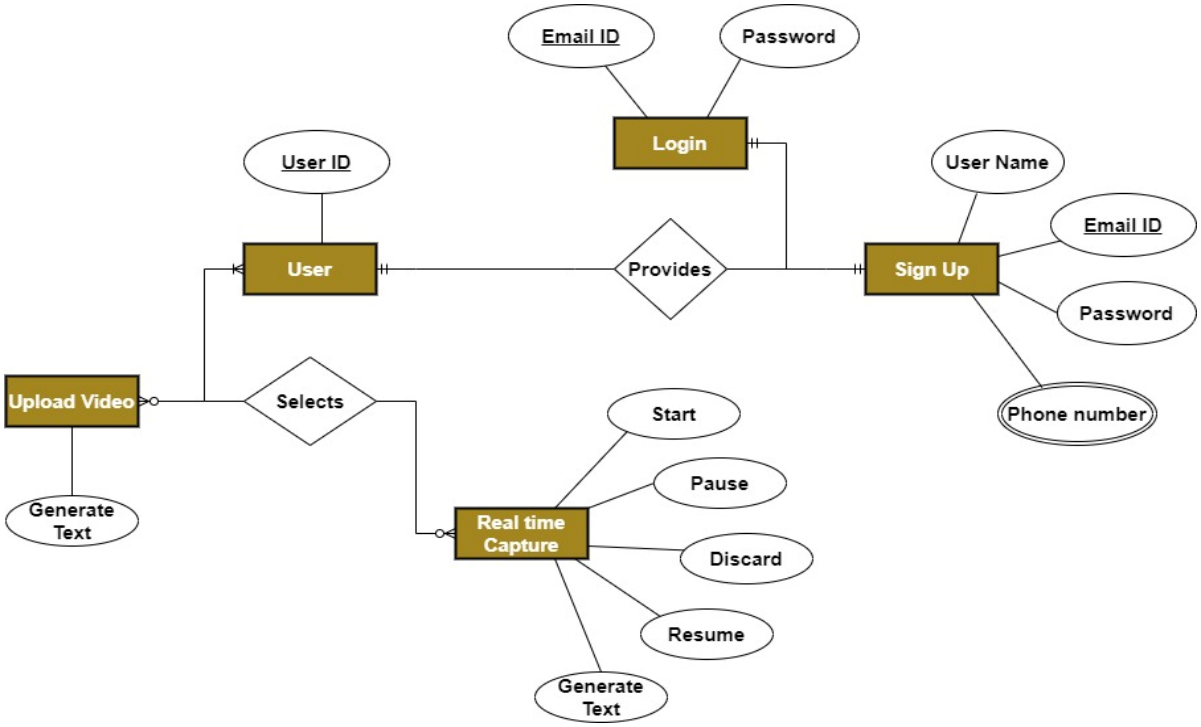

Figure 8: ER Diagram
This figure represents the ER diagram of our system

4.9.2 Data Dictionary

The user's data with their information provided while registration such as their name, email address, phone number and password with their unique user ID have been saved in the data dictionary for later login into the system.

Table 2: User's Data

This table is representing the information of user's data.

Field Name	Data Type	Size	P/F key	Example
User ID	Int	8	P	23
Name	Varchar	64		Raida Asma
Email Address	Varchar	64	P	raf@gmail.com
Phone Number	Varchar	12		0300-4072113
Password	Password	16		!#123abc@!

4.10 Risk Analysis

The risks that may be encountered during the project are as:

- The risk of user's confidential information such as the videos recorded on real time shall only be accessible with the private user only. It can be maintained with the help of open design principle.
- System shall not work out of the scope of words such as any word on which the data is not trained will consider an unknown word whose text will not be generated.

4.11 Conclusion

This chapter consists of functional, non-functional and other software and hardware requirements needed for the users and developers. Moreover, the detailed description of use cases with their basic and alternative flow have been discussed. In addition to this, the overview of the design has been explained with the ER diagram and the risks which can be encountered in this project.

Chapter 5: Proposed Approach and Methodology

There are different approaches used for translating the sign language into text but the studies showed that the suitable approach for converting the isolated videos into text is media pipe holistic with the LSTM layers. Our proposed approach explanation is being discussed below.

5.1 Media Pipe Holistic

We have first made the media pipe model accessible. Next, we took the following actions.

1) Extracting Key Points:

We took out our webcam. After starting a loop over every frame, we read the frames. These images move so quickly that they resemble video capture. We take that BGR-formatted image after receiving the OpenCV stream. However, we need to transform that picture from BGR to RGB in order to use it for detection. The next step is detection. For this, the flag is initially set to false to make the RGB image unwritable, then after making predictions, it is then set to true to make the image writable again. Then convert the pictures back to BGR format. We have created stylized landmarks utilizing the detection findings and the output image they produced. To set them apart from one another, these landmarks are essentially colorful landmarks. Then, as seen in the following figure, these historic photos are displayed on the screen:

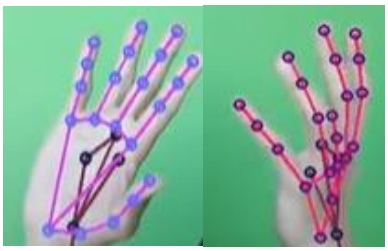

Figure 9: Landmarks
This figure represents the colourful landmarks

The relevant information for each body component is then retrieved and entered into a numpy array. These points are all combined into one numpy array. The necessary data is subsequently gathered in the folders for detection.

2) Key Points collection for training and testing:

After that we collect key points of videos for training purposes.

5.2 Data Pre-processing

Using a built-in library (train_test_split), we divided the data into two groups during data pre-processing, i.e. test and training data. Each data sequence has labels that we have assigned to them. The binary version of these labels was then transformed.

5.3 LSTM and Dense layer

For training a deep neural network we added 3 LSTM layers and 3 dense layers into our network. Each layer consists of different numbers of units and the activation function used is Relu except for the last layer which uses Softmax activation function.

5.4 Perform real time sign language detection using OpenCV

To obtain a live webcam stream and conduct real-time operations on the feed, OpenCV, a Python library, is utilized. To make the video stream user-friendly, many cv2 features are employed, and expected output is shown on the cv2 live feed.

5.5 Conclusion

The methodology used is extracting the landmarks with media pipes and labeling the gestures by training dataset using deep learning neural network LSTM and dense layers after pre-processing the video and changing it into image frames.

Chapter 6: High-Level and Low-Level Design

The system design, constraints, development methods, architecture design, sequence diagrams of every use case and class diagram are being discussed in detail to understand the high and low level design of the system.

6.1 System Overview

The gesture decode system aims to terminate the communication barrier between normal and hearing impair people. The patients will be the user who can login to the system and generate text by recording their gestures. Their privacy will be maintained with the open close principle and the interface inversion principle in which the interfaces will be separated for each functionality and deep learning approaches such as LSTM will be used for generating text from sign language video. The main approach used for this system will be inside-out TDD.

6.2 Design Considerations

This section describes many of the issues which need to be addressed or resolved before attempting to devise a complete design solution.

6.2.1 Assumptions and Dependencies

The assumptions and dependencies regarding the system and its use are as:

- **Hardware Requirements:** It is compulsory for the system to have the camera if the user has to capture the real time video of words and should have enough space for the videos stored in the database.
- **Operating systems:** The operating system required for this system is either windows or android to run the exe file.
- **End-user characteristics:** User shall know the Urdu sign language and must have a knowledge of English to efficiently use the system.
- **Possible and/or probable changes in functionality:** This system converts the sign language into text. One of the possible changes in functionality is to further convert the text into voice to have a better clarification for the illiterate people. Moreover, its scope can be increased.

6.2.2 General Constraints

The global limitations or constraints that have a significant impact on the design of our system's software are given below:

- **Hardware or software environment:**

  Hardware must be over 8 or more GB RAM to store videos and capable of rendering graphics to capture the video of the user. Moreover, a system to run the exe file containing python libraries

- **End-user environment:**

  The equipment used must have a good quality camera to record the video of the gestures

- **Availability or volatility of resources:**

The user must have the sufficient space to store the videos to use the option ‘upload video from database’ for generating text.

- **Data repository and distribution requirements:**

  Python SQL will be used to store the database of the videos and the user’s information who have registered and made an account.

- **Security requirements:**

  User authorization should be performed to access their private accounts. Secure protocols will be used by the system to perform actions.

- **Memory and other capacity limitations**:

  Other than the training dataset, there are 200 videos for human diseases and 73 videos for human anatomy which will be stored in the database. Thus, to use them or upload them from the database to generate text, users should have sufficient memory.

- **Performance requirements:**

  The system shall respond quickly without any interruption with maximum accuracy.

- **Verification and validation requirement:**

  The users must be verified by logging into the system using their own credentials. No one can access another person’s account.

- **Language Constraints:**

  The users who know the Urdu sign language can use the system.

6.2.3 Goals and Guidelines

The main goal of this project is to provide an ease of access to the users for easily using the system. The interface provided is very simple, understandable and meaningful. The help and proper guidelines will be provided by maintaining the principle ‘keep it simple stupid’. It moreover emphasizes on the quick response of the system such as to generate text without any interruption and the user can record a video. The memory used should be sufficient to store the video.

6.2.4 Development Methods

The development method suitable for this project is agile methodology which is used to reduce the risks and enhance the performance with each increment by testing each module. It is beneficial to divide the work into subsections and achieve the specified goal within the time frame. Moreover, it also gives an opportunity to increase the performance and make suitable changes with every increment to achieve the maximum accuracy.

6.3 System Architecture

The system architecture consists of the following steps.

6.3.1 Video to Text Model

This module is a crucial part of the entire project and is in charge of carrying out all the labor-intensive tasks.

6.3.1.1 Video to Image Frames

This module accepts video data and converts it into image frames with pre-processing techniques.

6.3.1.2 Image to Text Generator

This module accepts image data in frames form and generates text.

6.3.2 System Platform

This concept aims to give patients a user-friendly system interface so they can communicate with the entire system.

6.3.2.1 Frontend

This is the interface that is shown to patients to interact with the system for capturing videos and other interactions with the system.

6.3.2.2 Backend

Between several backend services and the user's frontend interface, this module serves as a middleman. It is a method of connecting the actions the user is carrying out in their browser with the machine learning models that are accessible on the backend.

6.3.2.3 Access Manager

This module uses authentication in order to make sure that a particular user can have access to only his account. The user cannot have authority to access someone else's account. But all the users of our system can have access to videos present in the database of our system.

6.3.3 Data store

The data store is used to store different stores of data related to users that are using the program and our data store will also have some videos of some common gestures. We are using a Python SQL database because it provides good performance for our use case.

6.3.4 System Architecture Diagram

The above mentioned steps have been explained with the help of following architecture diagram.

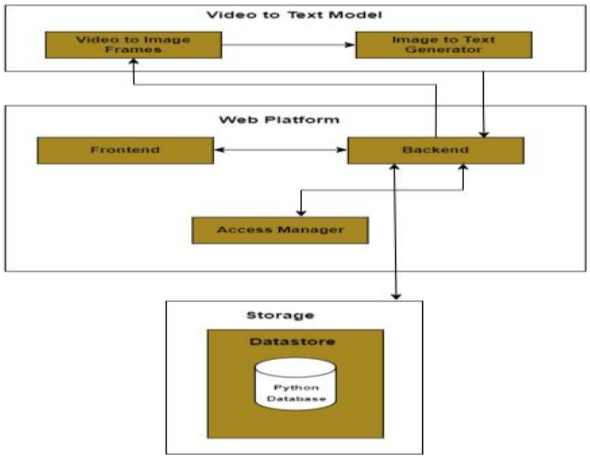

Figure 10: System Architecture Diagram

This figure represents the System Architecture Diagram

6.3.5 Subsystem Architecture

The subsystem architecture consists of the following steps:

6.3.5.1 Video Encoder

This module is responsible for converting the signs video into the image frames.

6.3.5.2 Image Frames

Image frames are generated and then these frames are pre-processed.

6.3.5.3 Feature Learning

The image frames are then added into the layers which checks the frames on the training dataset and predicts the output. It passes through the LSTM layer.

6.3.5.4 Feature Fusion and Classification

After passing through the fully connected layer, the features are being extracted and classified according to the classes and based on that it generates the text.

6.3.5.5 Subsystem Architecture Diagram

The figure below is describing the subsystem architecture diagram of the gesture decode system.

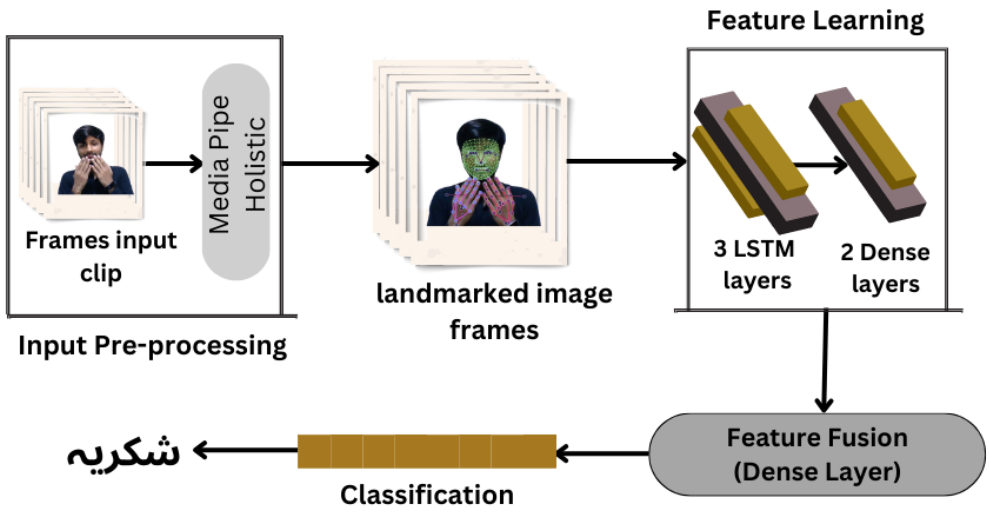

Figure 11: Subsystem Architecture Diagram
This figure represents the Subsystem Architecture Diagram

6.4 Architectural Strategies

The architectural strategies that we will be using for gesture decode system are given below:

6.4.1 Python Language

The system coding will be done in python language as it has many built in libraries such as OpenCV which can be used to convert the videos into image frames, Python is easy and understandable, easy language and preferred for deep learning. For implementing the frontend, GUI, react will be used.

6.4.2 Future Plans for Extension

We are initiating a work in Urdu sign language which is very vast. We defined the scope and limited it to some medical terms only. It can be extended by increasing the dataset such as covering more fields and shift it to the continuous sign language recognition for effective communication of hearing impair people.

6.4.3 Interface Paradigms

Interface used will be simple and easy so that it can be understood by every user. Every functionality defined will be meaningful.

6.4.4 Hardware and Software Interface Paradigms

To use the system, efficient memory is required to store the videos and a good quality camera should be needed to capture the clear videos to generate text.

6.4.5 Error Detection and Recovery

The error message will be displayed on the screen if the given input is wrong or out of the scope. Moreover, the user will be able to see if the generated text is not valid and can again capture the video and start over to translate the sign language.

6.4.6 Database Management

Microsoft SQL or mongo DB will be used to store the data of the users and the videos of the sign language.

6.5 Class Diagram

The overview of the system is given below as a class diagram of the system. It is explaining the dependency of the classes and their relationships. The functions and data members with their data type are being mentioned in each class.

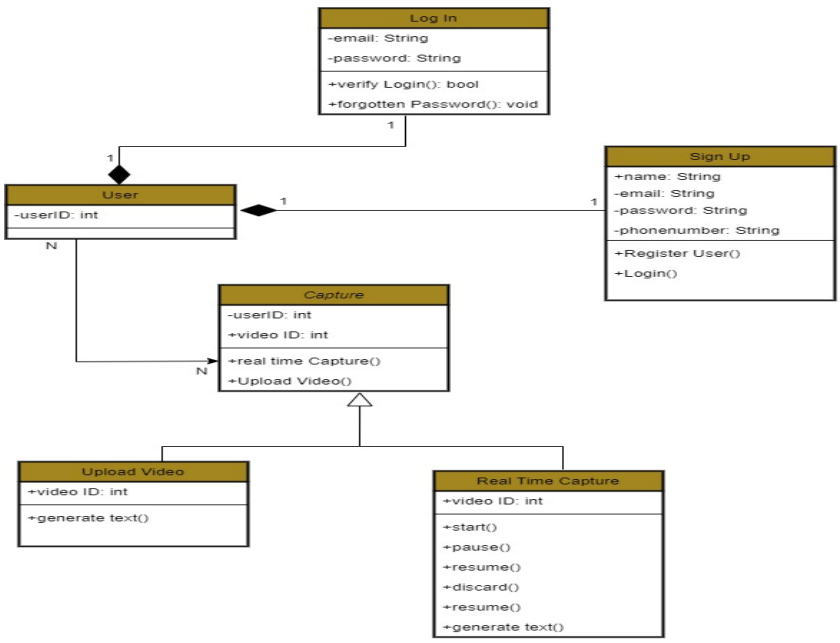

Figure 12: Class Diagram

This figure represents the Class Diagram of our system

6.6 Sequence Diagrams

The behavior design of the system is being defined through the sequence diagrams. In this section, the sequence diagram for every use case has been shown with detailed description of their work.

6.6.1 Signup

The new user will sign up into the system. For this purpose, he will enter all the required information which will be verified and stored in the database for later use.

If all the information is valid, the user will successfully be able to create an account and will be shown an error message otherwise.

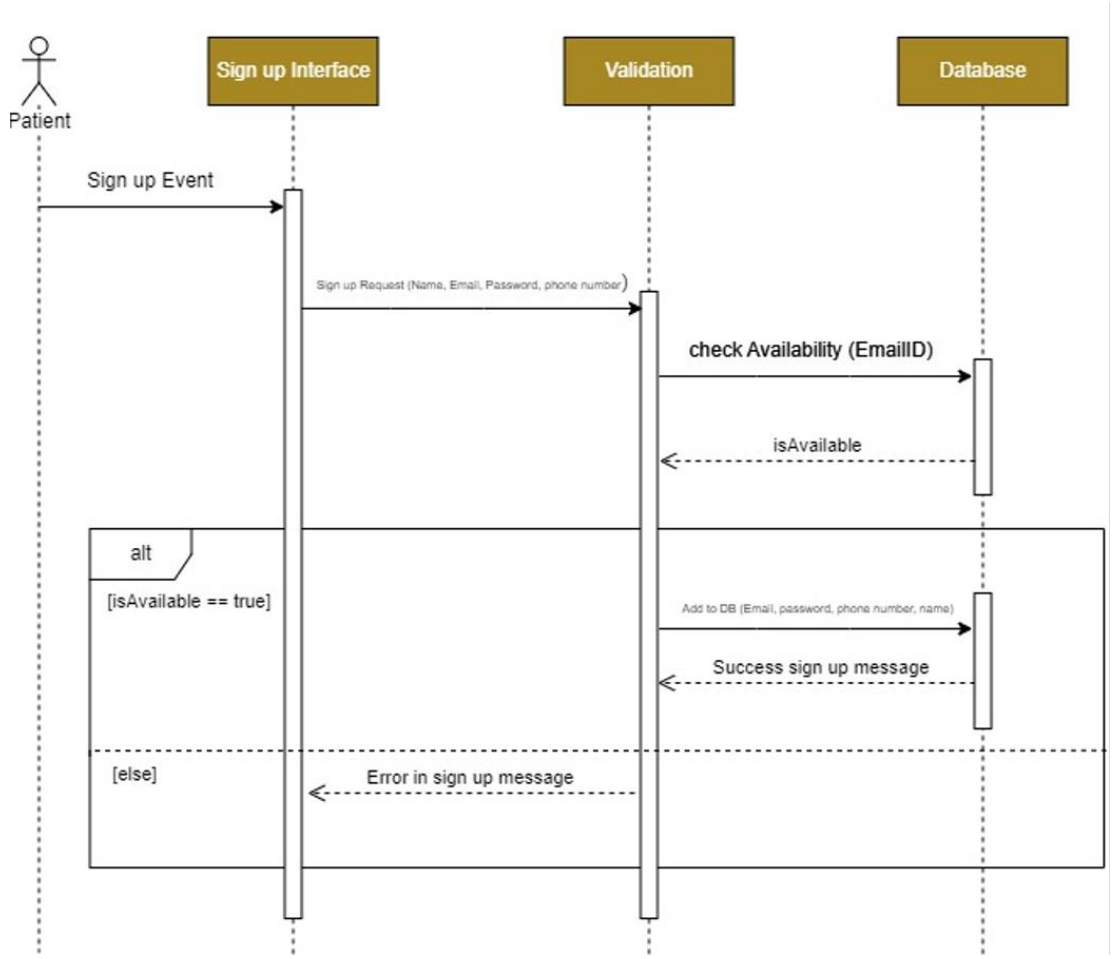

Figure 13: Signup

This figure represents the Signup Sequence Diagram of our system

6.6.2 Login

The user who has already registered for the system will be able to login to the system by entering the credentials.

The validity of the credentials will be verified every time by comparing them with the information saved in the database. If the provided information is correct, the user will successfully login to the system. Otherwise, an error message will be displayed.

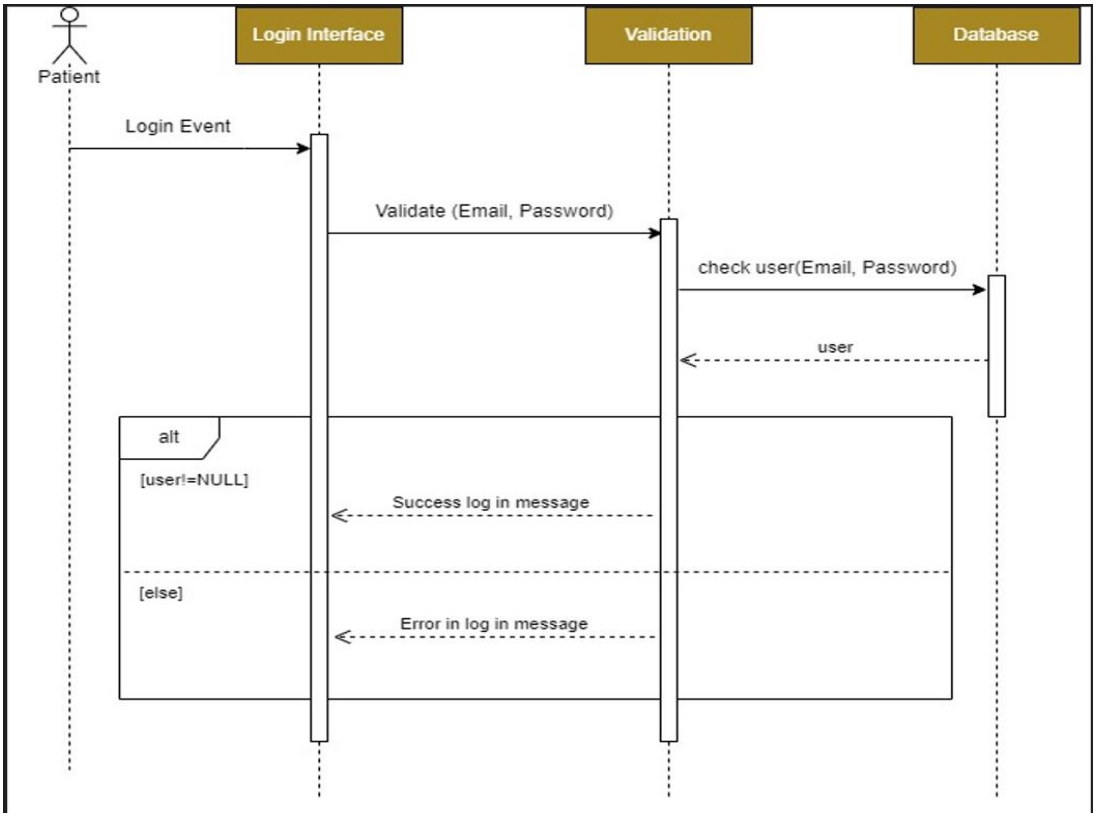

Figure 14: Login

This figure represents the Login Sequence Diagram of our system

6.6.3 Capture Button

When a user has to capture the video to generate the text of the sign language. He will choose the ‘capture’ on the home screen. When the request will be granted, the capture page will return the user with two options to record the video real time or to upload it from the database.

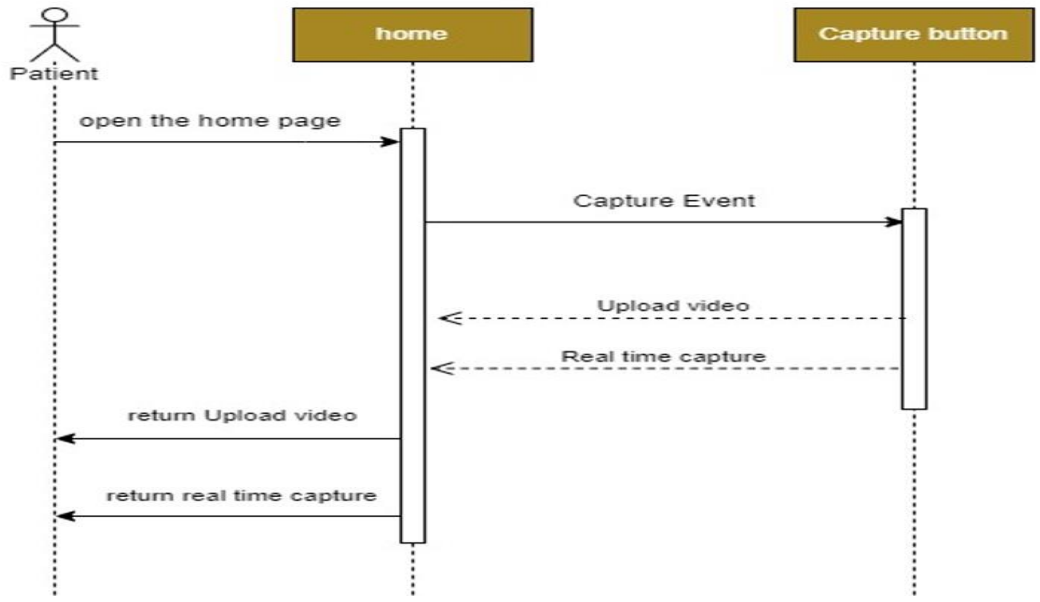

Figure 15: Capture Button

This figure represents the Capture Button Sequence Diagram of our system

6.6.4 Real Time Capture

After the capture button, if the user chooses to capture the video in real time, the system will ask the user to allow access to the camera for proceeding the procedure and will start capturing the user's video and the user can generate the text anytime.

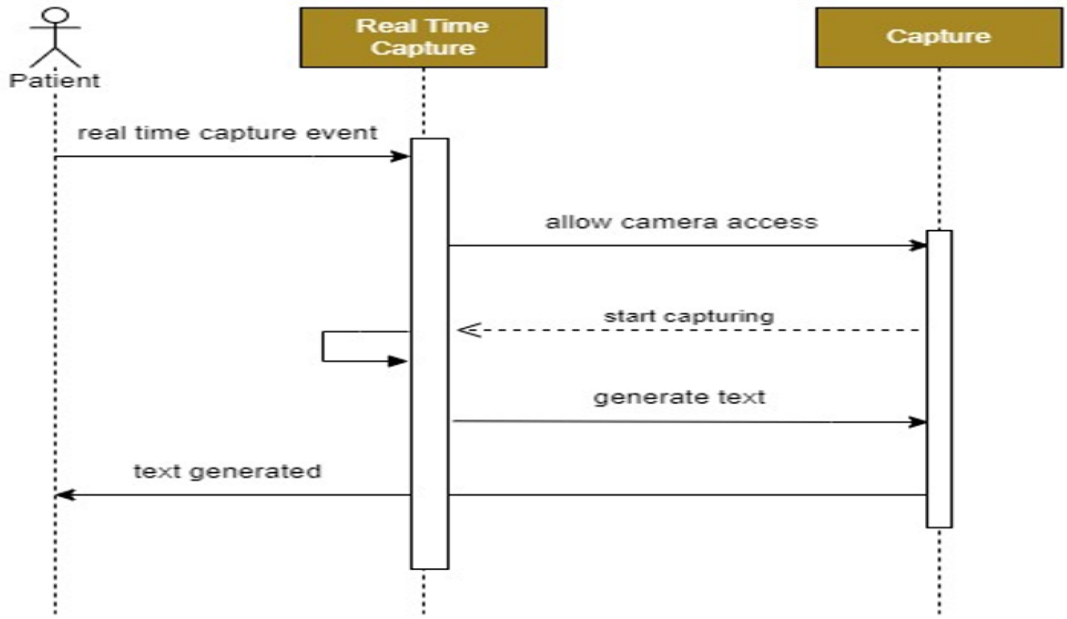

Figure 16: Real Time Capture

This figure represents the Real Time Capture Sequence Diagram of our system

6.6.5 Cancel the Sign

The user will be able to cancel/ discard the sign anytime during capturing the video by making a request to cancel the sign.

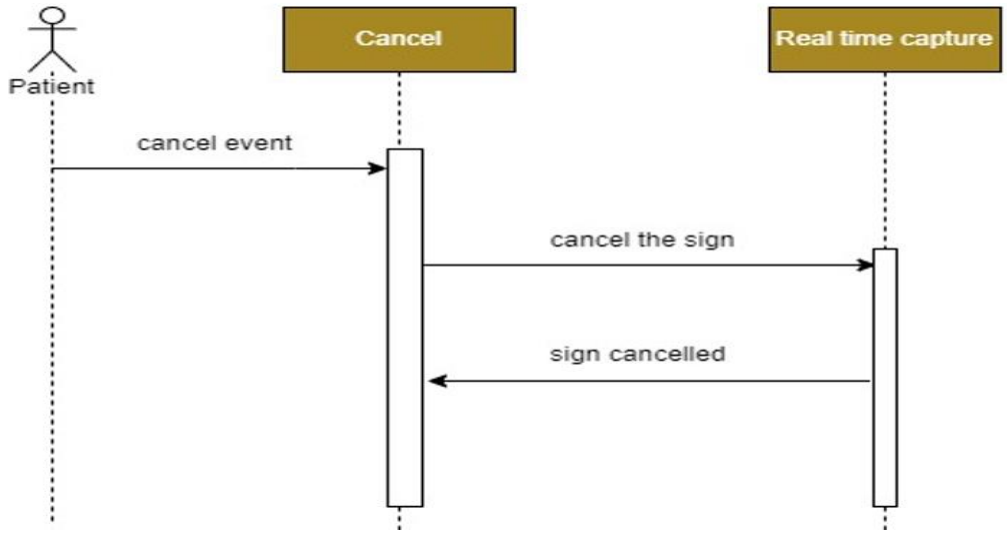

Figure 17: Cancel Button

This figure represents the Cancel Button Sequence Diagram of our system

6.6.6 Pause the Video

The user will be able to pause the video anytime during capturing the video by making a request to pause.

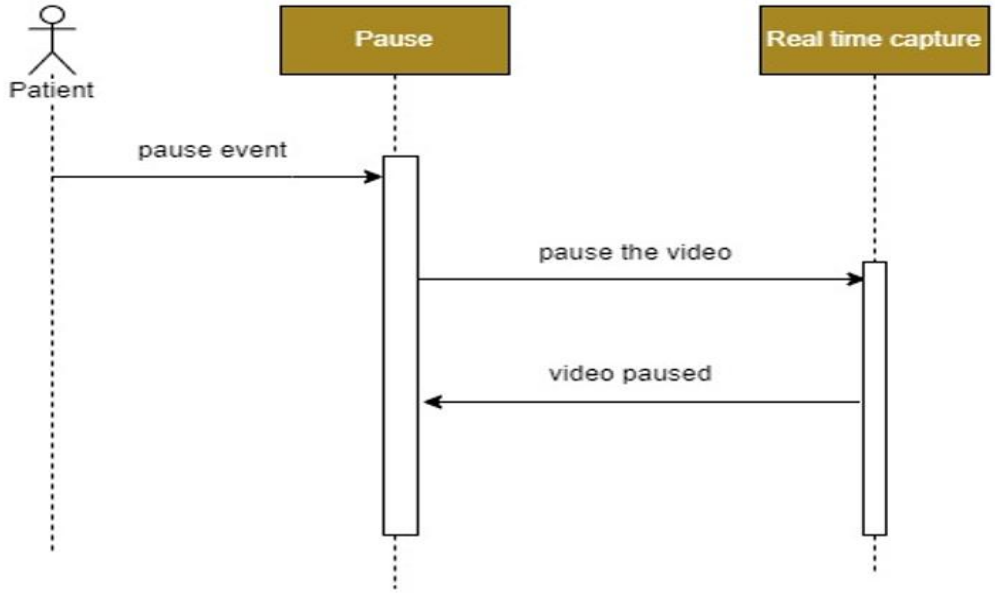

Figure 18: Pause Button
This figure represents the Pause the video Button Sequence Diagram of our system

6.6.7 Resume the Video

If the video is being paused, the user will be able to resume the video anytime during capturing the video by making a request to resume.

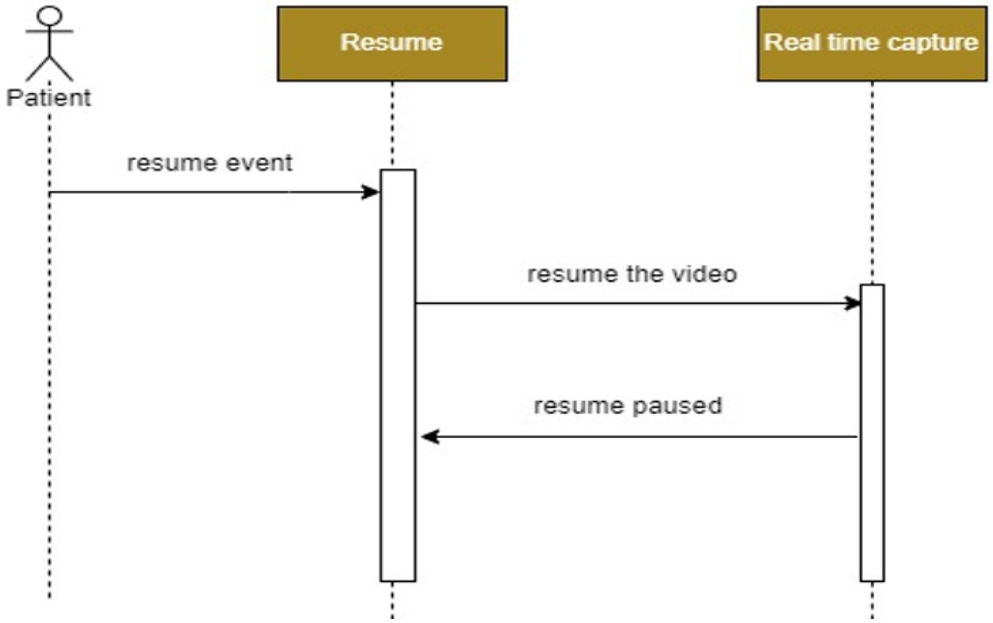

Figure 17: Resume Button
This figure represents the Resume the video Button Sequence Diagram of our system

6.7 Policies and Tactics

The planning and guidelines for policies and tactics that can affect the interface and implementation of our system are as:

6.7.1 Specific product to use

The software used for the project will be VS code as it is very fast and reliable software which is compatible for all the programming languages. It does not require too much memory space and works well for all languages by installing the extensions.

6.7.2 Coding Guidelines and conventions

The language preferred for the system is ‘python’ as it is easy to write, learn and understand. The code written will be easy, simple and full of comments to be understandable easily by everyone and will try to follow the solid principles.

6.7.3 Model use

The deep learning mechanisms give the best accuracy to translate the sign language. The proposed model is LSTM for training the model in which the videos will be divided into image frames and features will be extracted through media pipe holistic.

6.7.4 Testing the system

Black box testing strategy will be preferred for this project by following the unit testing and checking if the output is the desired one by giving the input such as we will learn the sign language words on which we have trained our system and will check if the generated text is valid.

6.7.5 Maintaining the system

The project is covering the medical domain and the system will be trained and tested on the particular words. Later, more dataset can be added.

6.8 Conclusion

The system architecture diagram is showing the processing of the super and sub view of the system, the class diagram tells us the relationships and functionalities of different classes and sequence diagrams display the detailed use of use cases. Moreover, constraints, architecture strategies and tactics clarify the language, model and product used for the system.

Chapter 7: Implementation and Test Cases

The main idea and procedures have been discussed in chapter 5 ‘methodology’. The main concern of this chapter is to describe the detailed work of the implementation of our project. The key points from the frames are being extracted and then trained with the collected dataset.

7.1 Implementation

We have used media pipe holistic and LSTM model in our implementation. Their description is given below:

7.1.1 Media Pipe Holistic

To extract key points from hands, body and face we are using a media pipe library. Media pipe extracts the key points from body parts and helps in detecting actions, pose etc. The results from using media pipe functions are then converted into numpy arrays and these arrays are used for further processing. So, we are using built- in library functions of the media pipe to extract key points from our video.

7.1.2 Neural Network

The Sequential model of keras tensor flow is used here so that we can add multiple layers of LSTM and Dense to train our model. Three LSTM layers are used here with the activation function Relu and different number of units in each layer. At the last LSTM layer the return sequence is set to false. The LSTM layers are followed by three dense layers with different number of units and activation function Relu in first two layers and Softmax in last layer to predict the output of probability sum 1. This architecture is used because we have very few datasets and we want our model to predict in real time. So, we used a combination of MP-holistic and LSTM layers to get fast results and good accuracy. The Sklearn Metrics library is used to calculate the confusion matrix.

7.1.3 OpenCV

The library of python named OpenCV is used to get live feed from webcam and perform real time operations on the video feed. Different functions of cv2 are used to make the video feed user friendly and predicted output is displayed on the cv2 live feed.

7.2 Test case Design and description

This section include input constraints that are true for every input in the set of associated test cases, any shared environmental needs, any shared special procedural requirements, and any shared case dependencies.

7.2.1 User Login Test case

Login Module			
1			
Test Case ID:	*1*	**QA Test Engineer:**	*Asma Ahmed*
Test case Version:	*1*	**Reviewed By:**	*Dr. Saira Karim*
Test Date:	-	**Use Case Reference(s):**	*Login use case*
Revision History:	*None*		

Objective	*To check if a user is able to successfully login in the system.*
Product/Ver/Module:	*Login Module of Gesture Decode System.*
Environment:	*The website is up and running and the system is online.*
Assumptions:	*Log in button is visible to the user.*
Pre-Requisite:	*The user has been registered into the system.*

Step No.	Execution description	Procedure result
1	*User clicks on the login button.*	*System displays the login page.*
2	*User enters his/her credentials and clicks the login button.*	*System directs the user to the home page.*

Comments: The test case passed. Our system is working according to our needs.

☒ *Passed* ☐ *Failed* ☐ *Not Executed*

7.2.2 Register Test case

Register Module			
1			
Test Case ID:	*2*	**QA Test Engineer:**	*Asma Ahmed*
Test case Version:	*1*	**Reviewed By:**	*Dr. Saira Karim*
Test Date:	-	**Use Case Reference(s):**	*Signup use case*
Revision History:	*None*		
Objective	*To check if a user is able to successfully register into the system and his data has been stored in the database.*		
Product/Ver/Module:	*Register Module of Gesture Decode System.*		
Environment:	*The website is up and running and the system is online.*		
Assumptions:	*The Sign Up button is visible to the user.*		
Pre-Requisite:	*Null*		

Step No.	Execution description	Procedure result
1	*User clicks on the sign up button.*	*System displays the sign up page.*
2	*User enters his/her credentials and clicks the sign up button.*	*System will register the user.*

Comments: The test case passed. Our system is working according to our needs.

☒ *Passed* ☐ *Failed* ☐ *Not Executed*

7.2.3 Home Button Test case

Home Page Module			
1			
Test Case ID:	*3*	**QA Test Engineer:**	*Asma Ahmed*
Test case Version:	*1*	**Reviewed By:**	*Dr. Saira Karim*
Test Date:	-	**Use Case Reference(s):**	*null*
Revision History:	*None*		
Objective	*To check if a user is able to successfully use the home page button.*		
Product/Ver/Module:	*Home Page Module of Gesture Decode System.*		
Environment:	*The website is up and running and the system is online.*		

Assumptions:	*Home page button is visible to the user.*	
Pre-Requisite:	*The user has been logged into the system.*	
Step No.	**Execution description**	**Procedure result**
1	*User clicks on the home page button.*	*System displays the home page.*
Comments: The test case passed. Our system is working according to our needs.		
☒*Passed* ☐*Failed* ☐*Not Executed*		

7.2.4 Capturing Video Button Test case

Capturing Video Module			
1			
Test Case ID:	*4*	**QA Test Engineer:**	*Asma Ahmed*
Test case Version:	*1*	**Reviewed By:**	*Dr. Saira Karim*
Test Date:	-	**Use Case Reference(s):**	*Real Time Capture use case*
Revision History:	*None*		
Objective	*To check if a user is successfully able to use the Capturing video button on the website.*		
Product/Ver/Module:	*Capturing Video Module of Gesture Decode System.*		
Environment:	*The website is up and running and the system is online.*		
Assumptions:	*Capturing video button is visible to the user.*		
Pre-Requisite:	*The user has been logged into the system.*		
Step No.	**Execution description**	**Procedure result**	
1	*User clicks on the capturing video button.*	*System displays the video capturing page.*	
2	*User allows the webcam to capture the video.*	*System enable webcam.*	
Comments: The test case passed. Our system is working according to our needs.			
☒*Passed* ☐*Failed* ☐*Not Executed*			

7.2.5 Start Capturing Button Test case

Capturing Video Module			
1			
Test Case ID:	*5*	**QA Test Engineer:**	*Raida Munir*
Test case Version:	*1*	**Reviewed By:**	*Dr. Saira Karim*
Test Date:	-	**Use Case Reference(s):**	*Real Time Capture use case*
Revision History:	*None*		
Objective	*To check if a user is able to start capturing video.*		
Product/Ver/Module:	*Capturing Video Module of Gesture Decode System.*		
Environment:	*The website is up and running and the system is online.*		
Assumptions:	*Start capturing button is visible to the user.*		
Pre-Requisite:	*The user is on the capturing video page and gives access to the webcam.*		
Step No.	**Execution description**	**Procedure result**	

1	*User clicks on the start capturing button.*	*System will start capturing the video.*
Comments: The test case passed. Our system is working according to our needs.		
☒*Passed* ☐*Failed* ☐*Not Executed*		

7.2.6 Pause Video Capturing Button Test case

Capturing Video Module			
1			
Test Case ID:	*6*	**QA Test Engineer:**	*Raida Munir*
Test case Version:	*1*	**Reviewed By:**	*Dr. Saira Karim*
Test Date:	-	**Use Case Reference(s):**	*Pause button use case*
Revision History:	*None*		
Objective	*To check if a user is able to pause capturing video.*		
Product/Ver/Module:	*Capturing Video Module of Gesture Decode System.*		
Environment:	*The website is up and running and the system is online.*		
Assumptions:	*Pause capturing button is visible to the user.*		
Pre-Requisite:	*The user is on the capturing video page and has started capturing the video.*		

Step No.	**Execution description**	**Procedure result**
1	*User clicks on the pause capturing button.*	*System will pause the video.*
Comments: The test case passed. Our system is working according to our needs.		
☒*Passed* ☐*Failed* ☐*Not Executed*		

7.2.7 Resume Video Capturing Button Test case

Capturing Video Module			
1			
Test Case ID:	*7*	**QA Test Engineer:**	*Raida Munir*
Test case Version:	*1*	**Reviewed By:**	*Dr. Saira Karim*
Test Date:	-	**Use Case Reference(s):**	*Resume button use case*
Revision History:	*None*		
Objective	*To check if a user is able to resume capturing the video.*		
Product/Ver/Module:	*Capturing Video Module of Gesture Decode System.*		
Environment:	*The website is up and running and the system is online.*		
Assumptions:	*Resume capturing button is visible to the user.*		
Pre-Requisite:	*The user is on the capturing video page and has paused capturing the video.*		

Step No.	**Execution description**	**Procedure result**
1	*User clicks on the resume capturing button.*	*System will resume the paused video.*
Comments: The test case passed. Our system is working according to our needs.		
☒*Passed* ☐*Failed* ☐*Not Executed*		

7.2.8 Discard Button Test case

Discard Button Module			
1			
Test Case ID:	*8*	**QA Test Engineer:**	*Raida Munir*
Test case Version:	*1*	**Reviewed By:**	*Dr. Saira Karim*
Test Date:	-	**Use Case Reference(s):**	*Cancel button use case*
Revision History:	*None*		
Objective	*To check if a user is able to discard the captured video.*		
Product/Ver/Module:	*Discard Button Module of Gesture Decode System.*		
Environment:	*The website is up and running and the system is online.*		
Assumptions:	*Discard button is visible to the user.*		
Pre-Requisite:	*The user is on the capturing video page and has started capturing the video.*		

Step No.	Execution description	Procedure result
1	*User clicks on the discard button.*	*System will discard the captured video.*
Comments: The test case passed. Our system is working according to our needs.		
☒*Passed* ☐*Failed* ☐*Not Executed*		

7.2.9 Stop Video Capturing Button Test case

Capturing Video Module			
1			
Test Case ID:	*9*	**QA Test Engineer:**	*Faseeh Ullah Jafar*
Test case Version:	*1*	**Reviewed By:**	*Dr. Saira Karim*
Test Date:	-	**Use Case Reference(s):**	*Stop button use case*
Revision History:	*None*		
Objective	*To check if a user is successfully able to stop capturing the video.*		
Product/Ver/Module:	*Capturing Video Module of Gesture Decode System.*		
Environment:	*The website is up and running and the system is online.*		
Assumptions:	*Stop button is visible to the user.*		
Pre-Requisite:	*The user is on the capturing video page and has started capturing the video.*		

Step No.	Execution description	Procedure result
1	*User clicks on the stop button.*	*System will stop capturing the video and the generated text button will be visible.*
Comments: The test case passed. Our system is working according to our needs.		
☒*Passed* ☐*Failed* ☐*Not Executed*		

7.2.10 Text Generation after Capturing Video Test case

Generate Text Module			
1			
Test Case ID:	*10*	**QA Test Engineer:**	*Faseeh Ullah Jafar*
Test case Version:	*1*	**Reviewed By:**	*Dr. Saira Karim*
Test Date:	-	**Use Case**	*Generate Text use case*

		Reference(s):	
Revision History:	*None*		
Objective	*To check if a user is successfully able to generate text of the captured video.*		
Product/Ver/Module:	*Generate text Module of Gesture Decode System.*		
Environment:	*The website is up and running and the system is online.*		
Assumptions:	*Generate text button is visible to the user.*		
Pre-Requisite:	*The user is on the capturing video page and has stopped capturing the video.*		
Step No.	**Execution description**	**Procedure result**	
1	*User clicks on the generate text button.*	*System displays the relevant text.*	
Comments: The test case passed. Our system is working according to our needs.			
☒*Passed* ☐*Failed* ☐*Not Executed*			

7.2.11 Upload Video Button Test case

Uploading Video Module			
1			
Test Case ID:	*11*	**QA Test Engineer:**	*Faseeh Ullah Jafar*
Test case Version:	*1*	**Reviewed By:**	*Dr. Saira Karim*
Test Date:	-	**Use Case Reference(s):**	*Upload Video Button use case*
Revision History:	*None*		
Objective	*To check if a user is successfully able to go onto the upload video page.*		
Product/Ver/Module:	*Uploading Video Module of Gesture Decode System.*		
Environment:	*The website is up and running and the system is online.*		
Assumptions:	*Upload video button is visible to the user.*		
Pre-Requisite:	*The user has been logged into the system.*		
Step No.	**Execution description**	**Procedure result**	
1	*User clicks on the uploading video button.*	*System displays the uploading video page.*	
Comments: The test case passed. Our system is working according to our needs.			
☒*Passed* ☐*Failed* ☐*Not Executed*			

7.2.12 Upload Video from Database Test case

Uploading Video Module			
1			
Test Case ID:	*12*	**QA Test Engineer:**	*Faseeh Ullah Jafar*
Test case Version:	*1*	**Reviewed By:**	*Dr. Saira Karim*
Test Date:	-	**Use Case Reference(s):**	*Upload Video button use case*
Revision History:	*None*		
Objective	*To check if a user is successfully able to upload the video from the stored videos in the database.*		
Product/Ver/Module:	*Uploading Video Module of Gesture Decode System.*		

Environment:	*The website is up and running and the system is online.*	
Assumptions:	*Upload video button is visible to the user.*	
Pre-Requisite:	*The user is on the uploading video page.*	
Step No.	**Execution description**	**Procedure result**
1	*User clicks on the uploading video button.*	*System displays the video uploading option.*
2	*User selects the video to upload.*	*System uploads the video.*
Comments: The test case passed. Our system is working according to our needs.		
☒ *Passed* ☐ *Failed* ☐ *Not Executed*		

7.2.13 Text Generation after Uploading Video Test case

Generate Text Module			
1			
Test Case ID:	*13*	**QA Test Engineer:**	*Faseeh Ullah Jafar*
Test case Version:	*1*	**Reviewed By:**	*Dr. Saira Karim*
Test Date:	-	**Use Case Reference(s):**	*Generate Text use case*
Revision History:	*None*		
Objective	*To check if a user is successfully able to generate text by uploading the video.*		
Product/Ver/Module:	*Generating text Module of Gesture Decode System.*		
Environment:	*The website is up and running and the system is online.*		
Assumptions:	*Generate text button is visible to the user.*		
Pre-Requisite:	*The user is on the uploading video page and uploaded a video.*		
Step No.	**Execution description**	**Procedure result**	
1	*User clicks on the generate text button.*	*System displays the relevant text.*	
Comments: The test case passed. Our system is working according to our needs.			
☒ *Passed* ☐ *Failed* ☐ *Not Executed*			

7.3 Test Metrics

Following the test metrics and their results after the testing process.

7.3.1 Test cases Matric

Metric	Purpose
Number of Test Cases	13
Number of Test Cases Passed	13
Number of Test Cases Failed	0
Test Case Defect Density	0/13
Test Case Effectiveness	13/13
Traceability Matrix	See Appendix A

7.4 Conclusion

The project consists of the videos captured in real time with the help of pre-processing techniques, it extracts the image frames and the points from each image frame are extracted through media pipes. Then, the data is being trained using the LSTM neural network. Finally, the gestures will be labeled and searched in the file to generate the text.

Chapter 8: User Manual

The user manual is given below:

8.1 Introduction

Gesture decode system comprises translating the Urdu Sign Language into Urdu text and helps the hearing impaired people to communicate effectively:

8.1.1 Overview of Gesture Decode System

The Gesture Decode System is a project that utilizes technology to translate Urdu Sign Language gestures into Urdu text. It employs gesture recognition algorithms and linguistic rules to accurately interpret sign language gestures and generate corresponding text output.

8.1.2 Purpose and Benefits of Gesture Decode System

The purpose of the Gesture Decode System is to facilitate communication for individuals who are deaf or hard of hearing and use Urdu Sign Language as their primary means of communication. The system enables them to express themselves in written Urdu text, which can be easily understood by others. The benefits of the system include improved communication, increased accessibility, and enhanced social integration for users who rely on sign language for communication.

8.1.3 System Requirements and Compatibility

The Gesture Decode System may require specific hardware and software requirements for optimal performance. These requirements may include a camera or sensor for gesture recognition, a device with sufficient processing power, and compatible operating systems or platforms. The system's compatibility may vary depending on the devices or platforms it is designed to work with, and it may require regular updates or maintenance to ensure compatibility with the latest technologies or software updates.

8.2 Getting Started

In order to start Gesture decode system, these steps need to be followed:

8.2.1 Installation and Setup Instructions

It provides step-by-step instructions on how to install and set up the Gesture Decode System on a user's device. The user needs to download and install the necessary software, connecting the camera or sensor, and configuring any settings or preferences.

8.2.2 System Configuration and Calibration

It specifies how users can configure and calibrate the system for optimal performance. This includes adjusting camera settings, gesture recognition sensitivity, or other system parameters to ensure accurate gesture recognition and text translation.

8.2.3 User Registration and Login process

Provide instructions on how users can register an account and log in to the Gesture Decode System. A user has to create a profile, setting up authentication, and managing user credentials for accessing the system's features and functionalities.

8.3 System Interface

Gesture decode system provides a user friendly interface and performs the actions as per requirement.

8.3.1 System Navigation and User Interface Overview

The system is implemented using React and JavaScript. A user is able to read the instructions from the about page whenever he feels any ambiguity in using the system. After signing in, he will be able to see two options: real time capturing and translation through video uploading.

8.3.2 Real Time Gesture Recognition Feature

In this feature, a user will be able to see the portion of video capturing and a button to start capturing the video. After pressing the start button, pause, resume and stop buttons will be enabled. Once the video is captured and stopped. User needs to demonstrate proper gestures for accurate recognition.

8.3.3 Gesture Recognition through Video Uploading

In this feature, a button to upload video from the database is mentioned on the screen and the user can choose any sign language video from the dataset.

8.3.4 How to use the Text Translation Feature

In both the features mentioned above, a text generation button will be enabled, when the user will stop the video in real time capturing or will upload the video from the database. After pressing this button, a text of the particular sign will be shown on the screen.

8.4 Gesture Recognition

The details of Gesture Recognition are as follows:

8.4.1 User Requirement for Gesture Recognition

Users need to show the half upper part of their body in the video capture. Emphasize the importance of capturing clear and accurate signs to ensure accurate gesture recognition.

8.4.2 Extracting Landmarks from Image Frames

System generates image frames from the video to extract landmarks on the user's body, such as hand movements, facial expressions, and body postures which are used to identify the sign language gestures.

8.4.3 Importance of Lighting and Distance

User needs to maintain an appropriate distance from the camera for accurate gesture recognition. Poor lighting or being too far from the camera may affect the system's ability to accurately recognize sign language gestures.

8.4.4 Feedback on Gesture Recognition Accuracy

The system will generate the text and the user will be notified on successful or unsuccessful gesture recognition, and suggestions on how to improve gesture recognition accuracy.

8.4.5 Troubleshooting and FAQs related to Gesture Recognition

Include a troubleshooting guide and frequently asked questions related to gesture recognition, along with solutions or workarounds for common issues or challenges users may encounter when performing sign language gestures for recognition.

8.5 System Settings

User need to understand the following settings:

8.5.1 Camera Settings

The system should be given the camera access and clear camera results are required for the Gesture Decode System to capture video properly and translate the sign. It includes accurately adjusting camera resolution, frame rate to fit, and focus to ensure clear and accurate video capture.

8.5.2 Video Input Selection

Provide instructions on how users can select the appropriate video input for the Gesture Decode System, whether it's from the real-time video capture or from uploaded videos from a dataset. Also, how to switch between different video sources and how to ensure the selected video input is compatible with the system.

8.5.3 Saving and Loading Settings

Provide instructions on how users can save and load system settings, if applicable or required. A user can save custom configurations for future use or can load the predefined settings for different scenarios or environments.

8.6 Maintenance and Troubleshooting

It is important to provide comprehensive maintenance and troubleshooting information, including clear instructions, contact information for technical support, and FAQs, to assist users in resolving issues and ensuring the proper functioning of the Gesture Decode System:

8.6.1 Regular System Maintenance Guidelines

The user should take care of the following things: cleaning the camera lens, checking for software updates, and ensuring proper hardware functioning, if applicable. Emphasize the importance of regular maintenance to avoid potential issues and ensure optimal performance.

8.6.2 Troubleshooting Common Issues

It encounters the proper troubleshooting guide that addresses common issues users may face such as video capture problems, gesture recognition accuracy, or system configuration errors.

8.6.3 Frequently Ask Questions (FAQs)

It provides a list of frequently asked questions related to the Gesture Decode System. Cover common inquiries, such as system compatibility, installation process, gesture recognition accuracy, and other relevant topics. Provide clear and concise answers to help users quickly find solutions to their queries.

8.7 Conclusion

This user manual has provided an overview of the Gesture Decode System, including its purpose, benefits, system requirements, and compatibility. It has also covered important aspects such as getting started with installation and setup instructions, system configuration and calibration, user registration and login process, and system settings.

Additionally, maintenance and troubleshooting guidelines, along with common issue resolutions and contact information for technical support, have been provided to assist users in ensuring the smooth operation of the system. The inclusion of frequently asked questions (FAQs) helps users quickly find solutions to common queries.

By following the instructions and guidelines provided in this user manual, users can effectively utilize the Gesture Decode System to communicate in Urdu text using Urdu Sign Language, enhancing their ability to interact and connect with others.

Note: It is important to provide clear instructions and guidelines on proper sign language techniques and the system's requirements for accurate gesture recognition to ensure users can effectively use the Gesture Decode System for translation.

Chapter 9: Experimental Results and Discussion

In this project, we have made a prototype and trained it with the dataset collected from schools in Urdu sign language. We have tested and analyzed the results of five isolated videos by training them on the dataset. We have taken five videos for each word to train.

As shown in the table, the average accuracy is 54%. Similar words are not being detected correctly because of the small dataset.

Table 3: Accuracy of Videos

The accuracies of different words for sign language recognition are presented here.

Words	**Training Videos**	**Accuracy**
Ankle	35	57%
Blood	35	46%
Eye	35	54%
Fist	35	45%
Heart	35	56%
Jaw	35	47%
Knuckle	35	61%
Lips	35	64%
Palm	35	58%
Skull	35	49%
Thumb	35	68%

9.1 Conclusion

We have trained our dataset on different videos but due to the deficiency in the dataset, the accuracy is 39% overall which is very low. To improve this, we will collect more dataset and will apply the augmentation techniques to increase the dataset and enhance the accuracy.

Chapter 10: Conclusion and Future Work

Communication is the basic need to follow up in society. If you are not able to convey your message then it's not easy for you to be successful. Some people are introverts which can practice in different activities to enhance their communication skills but there are some people who use different languages to communicate which a normal person cannot understand. Thus, it becomes the need to overcome the communication gap for the hearing impaired people. In this project, we are building an idea to translate the hearing impaired people language into understandable text. This will lead them to actively participate in day to day activities. Various sign language translations in different languages have been done but no such work has been done for Urdu sign language. The main reason could be the deficiency of the dataset available which we are generating by going to hearing impaired schools and colleges, through augmentation techniques and some self-made videos by learning sign language.

Our project is a research and development based project that essentially solves the problem of patients who cannot speak or hear. This document begins with an overview and introduction to the project. It consists of class diagrams, architecture diagrams and sequence diagrams which will help you understand the methodology of this project. These patients were then investigated in this phase of the project, that is, we collected various videos of these patients and trained the dataset. Additionally, we learned about various backend technologies and the python APIs, libraries and have successfully used different techniques to generate the dataset and methods to improve the accuracy. Detailed usage examples for all actors are provided in this document. A basic entity-relationship model for the database is also provided. The documentation is complete and the initial development of the project on this report has been completed.

The next stage is mainly the development stage, in which we will be collecting more videos and we will train our dataset on a large scale. Main focus would be to increase the accuracy of our system. This is how our application works and relieves deaf-mutes.

Our system, gesture decode system, is covering the domain of human anatomy and diseases based on isolated videos in Urdu language. It can be enhanced by training the model to generate the sentence to sentence continuous sign language translation on a large scale by increasing the scope.

References

[1] R. Harini, R. Janani, S. Keerthana, S. Madhubala and S. Venkatasubramanian. (2020). "Sign Language Translation." *6th International Conference on Advanced Computing and Communication Systems* [Online]. pp 883-886. Available: https://ieeexplore.ieee.org/abstract/document/9074370

[2] Necati Cihan Camgoz, Oscar Koller, Simon Hadfield, Richard Bowden. (2020). "Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation." *IEEE/CVF Conference on Computer Vision and Pattern Recognition* [Online]. pp 10023-10033. Available: https://openaccess.thecvf.com/content_CVPR_2020/html/Camgoz_Sign_Language_Transformers_Joint_End-to End_Sign_Language_Recognition_and_Translation_CVPR_2020_paper.html

[3] Gautham. Jayadeep, N. V. Vishnupriya, Vyshnavi. Venugopal, S. Vishnu, M. Geetha. (2020). "Mudra: Convolutional Neural Network based Indian Sign Language Translator for Banks." *4th International Conference on International Computing and Control Systems* [Online]. pp 1228-1232. Available: https://ieeexplore.ieee.org/abstract/document/9121144

[4] Kayo. Yin, Jesse. Read. (2020, December). "Better Sign Language Translation with STMC-Transformer." *28th International Conference on Computational Linguistics* [Online]. pp 5975-5989. Available: https://aclanthology.org/2020.coling-main.525/?ref=https://githubhelp.com

[5] Neena Aloysius, M. Geetha. (2020, May 17). "Understanding vision-based continuous sign language recognition." *Springer Science Business Media, LLC, part of Springer Nature* [Online]. pp 22177–22209. Available: https://fci.stafpu.bu.edu.eg/Computer%20Science/1273/publications/nada%20bahaa%20ibrahim%20ahmed_1205-1227.pdf

[6] K. Tiku, J. Maloo, A. Ramesh and I. R. (2020, December). "Real-time Conversion of Sign Language to Text and Speech." *Second International Conference on Inventive Research in Computing Applications* [Online]. pp 346-351. Available: https://ieeexplore.ieee.org/abstract/document/9182877

[7] M. Al-Hammadi *et al*. (2020, December). "Deep Learning-Based Approach for Sign Language Gesture Recognition with Efficient Hand Gesture Representation." *IEEE Access* [Online]. vol. 8, pp 192527-192542. Available: https://ieeexplore.ieee.org/abstract/document/9229417

[8] Z. Zhou, V. W. L. Tam and E. Y. Lam. (2021). "Sign BERT: A BERT-Based Deep Learning Framework for Continuous Sign Language Recognition." *IEEE Access* [online], vol. 9 Available: https://ieeexplore.ieee.org/document/9635818

[9] Ilias Papastratis, Kosmas Dimitropoulos and Petros Daras. (2021). "Continuous Sign Language Recognition through a Context-Aware Generative Adversarial." *Network* [Online]. Available: https://www.mdpi.com/1424-8220/21/7/2437#cite

[10] Deep Kothadiya, Chintan Bhatt, Krenil Sapariya, Kevin Patel, Ana-Belén Gil-González and Juan M. Corchado. (2022, Jun 03). "Deepsign Sign Language Detection and Recognition Using Deep Learning." [Online]. Available: https://www.mdpi.com/2079-9292/11/11/1780

[11] Maher Jebali, Abdesselem Dakhli and Mohammed Jemni. (2021). “Vision-based continuous sign language recognition using multimodal sensor fusion.” [Online]. Available: https://doi.org/10.1007/s12530-020-09365-y

[12] Yanliang Jin, Xiaowei Wu, LAN Ni and Xiaoqi Yu. (2022). “Continuous Sign Language Recognition Using Multiple Feature Points.” [Online]. Available: https://ieeexplore.ieee.org/document/9788698

[13] Xie, Pan and Cui, Zhi and Du, Yao and Zhao, Mengyi and Cui, Jianwei and Wang, Bin and Hu, Xiaohui. (2021). “Multi-Scale Local-Temporal Similarity Fusion for Continuous Sign Language Recognition.” [Online]. Available: https://arxiv.org/abs/2107.12762

[14] Y. Chen, X. Mei, X. Qin. (2022, August 24). “Two-stream lightweight sign language transformer.” *Machine Vision and Applications* 33 [Online], No. 79. Available:

https://link.springer.com/article/10.1007/s00138-022-01330-w

[15] R. Li, L. Meng. (2022, July 05). “Sign language recognition and translation network based on multi-view data.” *Applied Intelligence* [Online]. Vol. 52. Available: https://link.springer.com/article/10.1007/s10489-022-03407-5

[16] R. Lu, Q. Song. (2022, July 22). “Research on the improved gesture tracking algorithm in sign language synthesis.” *J Supercomputing* (2022). Available [Online]:

https://link.springer.com/article/10.1007/s11227-022-04705-y

[17] Boukdir, A., Benaddy, M., Ellahyani, A. *et al.* (2021, September 16). “Isolated Video-Based Arabic Sign Language Recognition Using Convolutional and Recursive Neural Networks”. *Arabian Journal for Science and Engineering* [Online]. Vol. 47. Available:

https://link.springer.com/article/10.1007/s13369-021-06167-5#author-information

[18] Y. Chen, F. Wei, X. Sun, Z. Wu, S. Lin. (2022). “A Simple Multi-Modality Transfer Learning Baseline for Sign Language Translation.” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR) [Online]. pp. 5120-5130, 2022. Available: https://openaccess.thecvf.com/content/CVPR2022/html/Chen_A_Simple_Multi-Modality_Transfer_Learning_Baseline_for_Sign_Language_Translation_CVPR_2022_paper.html

[19] S. Barathi, O. Bekhzod, N. Shraddha, K. Sangchul, P. Kil-Houm, K. Jeonghong. *Et al*. (2022, July 13). “An integrated media pipe-optimized GRU model for Indian sign language recognition.” *Sci Rep* 12 [Online]. 11964, 2022. Available: https://www.nature.com/articles/s41598-022-15998-7#citeas

[20] R. Cui, H. Liu and C. Zhang. (2019, January 01). “A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training.” in *IEEE Transactions on Multimedia* [Online]. Vol. 21, no. 7, pp. 1880-1891, July 2019. Available: https://ieeexplore.ieee.org/abstract/document/8598757/author

Appendix

Appendix A: Traceability Matrix

Any tables, figures, forms, or other materials that are not totally central to the analysis but that need to be included are placed in the Appendix.

Table 4: Traceability Matrix

This table shows the traceability matrix of Gesture Decode System

Sr.no	Requirements Description	Test Case ID	Test Case Status
1	Login to the website	1	Pass
2	Signup to the website	2	Pass
3	View Home Page	3	Pass
4	Press Video Capturing Button	4	Pass
5	Start Capturing Video	5	Pass
6	Pause Video Capturing	6	Pass
7	Resume Video Capturing	7	Pass
8	Discard an Sign	8	Pass
9	Stop Video Capturing	9	Pass
10	Generate Text after capturing the video	10	Pass
11	Press upload video from database button	11	Pass
12	Upload the video from database	12	Pass
13	Generate Text after uploading the video	13	Pass