REAL-TIME COMMUNICATION SYSTEM POWERED BY AI FOR SPECIALLY ABLED

ABSTRACT

People with impaired speech and hearing uses Sign language as a form of communication. Disabled People use this sign language gestures as a tool of non-verbal communication to express their own emotions and thoughts to other common people. Conversing with people having a hearing disability is a major challenge. Deaf and Mute people use hand gesture sign language to communicate, hence normal people face problems in recognizing their language by signs made. Hence there is a need for systems that recognize the different signs and conveys the information to normal people. But these common people find it difficult to understand their expression, thus trained sign language expertise are needed during medical and legal appointment, educational and training session. Over the past few years, there has been an increase in demand for these services. Other form of services such as video remote human interpret using the high-speed Internet connection, has been introduced, thus these services provides an easy to use sign language interpret service, which can be used and benefited, yet have major limitations. To address this problem, we can implement artificial intelligence technology to analyse the user's hand with finger detection. In this proposed system we can design the vision based system in real time environments. And then using deep learning algorithm named as Convolutional neural network algorithm to classify the sign and provide the label about recognized sign.

1. INTRODUCTION

1.1 PROJECT OVERVIEW

Sign language recognition is the process of translating the user's gestures and signs into text. It aids those who are unable to interact with the general population in communication Using image processing methods and neural networks, the motion is mapped to pertinent text in the training data, transforming unprocessed photos and videos into text that can be read and understood. People who are dumb are typically prohibited from having regular conversations with other people in society. They sometimes struggle to communicate with regular people through gestures because the majority of people only recognise a small number of them. People who are deaf or have hearing loss are unable to communicate vocally, so they must frequently use some type of visual communication. The primary form of communication for the deaf and dumb community is sign language. Similar to other languages, it contains grammar and vocabulary, but it communicates primarily through images.

1.2 PURPOSE

The problem occurs when those who are stupid or deaf attempt to use these grammars of sign language to interact with others. This is because the majority of individuals are not familiar with these grammar standards. It has been noted that a foolish person can only communicate with members of his or her family or the deaf community. The popularity of international programmes and the financing they get highlight the value of sign language. In this age of technology, a computer-based solution is highly desired by the dumb community. Teaching a computer to recognise speech, facial expressions of emotion, and human gestures are some steps toward achieving this goal. Gestures are used to convey information nonverbally. Humans are capable of an endless amount of motions at any given moment. Since human motions are seen visually, computer vision researchers are particularly interested in them. The project's objective is to develop an HCI that can recognise human motions. These motions must be translated into machine language using a challenging programming process. In our paper, we concentrate on Image Processing and Template Matching for better output creation.

2. LITERATURE REVIEW

2.1 EXISTING PROBLEM

2.1.1 TITLE: A STUDY ON ARABIC SIGN LANGUAGE RECOGNITION FOR DIFERENTLY ABLED USING ADVANCED MACHINE LEARNING CLASSIFERS

AUTHOR: MOHAMMED MUSTAFA,2021.

Around 70 million people use sign language worldwide, and an automated method for translating it could significantly improve communication between sign language users and those who might not understand it. Nonverbal communication that includes the use of other bodily parts is called sign language. Face expressions, together with movements of the hands, eyes, and lips used in sign language communication to communicate information. People who have trouble hearing or speaking rely heavily on sign language as a form of communication in daily life. The inconsistent shape, size, and posture of the hands or fingers in an image was however shown by computer translation of sign language, which was highly complicated. SLR can be used in two main ways: based on picture or sensor. The main advantage of image-based frameworks is that people do not need to use complicated equipment. In any case, the preprocessing process necessitates large computations. Sensors frameworks use gloves fitted with sensors rather of relying just on cameras. Like spoken language, sign language does not confined to a certain location or region. It is trained differently over the world (Shin et al. 2019). It is sometimes referred to as Chinese Sign Language, American Sign Language, African Sign Language, and Arabic Sign Language (ArSL). India does not have a standardised sign language with important modifications, unlike sign languages in Europe and America. However, a dictionary of ISL was just created by Coimbatore's Vivekananda University for the Ramakrishna Missions. there are nearly there are currently 2037 signs available in Indian Sign Language (ISL). Similar to how SLR models are separated into sensor glove based and vision-based categories. Recent research on SLR can be divided into contact-based and vision-based methods. Physical interaction between sensing devices is a component of the contact-based technique and clients. It often employs an instrumented glove that uses electromyography, inertial estimation, or electromagnetic to capture information on the executed sign's position, extension, direction, and angle.

2.1.2 TITLE: SIGN LANGUAGE TRANSFORMERS: JOINT END-TO-END SIGN LANGUAGE RECOGNITION AND TRANSLATION

AUTHOR: NECATI CIHAN CAMG¨OZ, 2021.

The translation is improved by having a mid-level sign gloss representation, which efficiently recognises the various signs, according to earlier research on sign language translation. Performance significantly In fact, gloss level tokenization is necessary for the state-of-the-art in translation to function. We present a unique architecture based on transformers that simultaneously learns Continuous Sign Language Recognition and Translation while being end-to-end trainable. This is accomplished by combining the recognition and translation issues into a single, unified architecture employing a Connectionist Temporal Classification (CTC) loss. This collaborative approach achieves significant performance improvements while simultaneously resolving two related sequence-to-sequence learning problems without the need for ground-truth timing information. The primary form of communication for the Deaf community is sign language, which is their native tongue. They use a variety of complementing channels as visual languages to communicate ideas. This comprises both manual and non-manual characteristics, such as head, shoulder, and torso movement as well as manual characteristics like hand shape, movement, and stance. The purpose of sign language translation is to either extract an equivalent spoken language sentence from written text or translate written text into a video of signs. A clip of someone doing the continuous sign. However, a large portion of this latter work is done in the field of computer vision, where linguists refer to these channels as articulators. Word embedding with spatial embedding has concentrated on understanding the order of sign glosses rather than providing a complete translation into a spoken language counterpart (Sign Language Translation, or SLT). This distinction is crucial because spoken and sign languages have significantly different grammatical structures. Word order variations, the use of multiple channels to convey simultaneous information, and the use of direction and space to indicate the relationships between objects are just a few examples of these differences.

2.1.3 TITLE: SIGN LANGUAGE RECOGNITION SYSTEMS: A DECADE SYSTEMATIC LITERATURE REVIEW

AUTHOR: ANKITA WADHAWAN,2020.

As spoken languages are pronounced with the lips and heard with the ear, they utilise the "vocal-auditory" channel. Additionally, all writing systems come from, or are spoken languages' representations. Because they use the "corporal-visual" channel, which is created with the body and perceived with the eyes, sign languages (SLs) are unique. SLs are widely used by the deaf communities but are not internationally recognized. They are considered natural languages because deaf people can spontaneously gather and communicate with one another anywhere. SLs have independent vocabularies and grammatical structures and are not descended from spoken languages . The signs that the deaf use actually have the same internal structure as spoken words. The signs of SLs are produced using a small number of different sounds, just as hundreds of thousands of English words are. A fixed number of gestural characteristics. As a result, signs are not complete gestures but rather can be analysed as a collection of linguistically important characteristics. A gloss, the basic component of an SL and the closest representation of a sign's meaning, is made up of combinations of the aforementioned qualities. SLs, comparable to the spoken ones contain a list of grammatically flexible rules that apply to both manual and non-manual elements. Signers utilise both of them concurrently (and frequently with a flexible temporal structure) to create phrases in an SL. A particular feature may be the most important consideration when interpreting a gloss, depending on the context. It can change a verb's meaning, provide spatial and temporal context, and distinguish between things and people. A signer's glosses can be inferred from video recordings using a process known as sign language recognition (SLR). Despite the fact that there is a lot of labour, There is a severe paucity of comprehensive experimental research in the subject of SLR. Additionally, most articles don't release their code or present findings from all available datasets. As a result, experimental findings in the field of SL are rarely repeatable and interpretable.

2.1.4 TITLE: A COMPREHENSIVE STUDY ON SIGN LANGUAGE RECOGNITION METHODS

AUTHOR: NIKOLAS ADALOGLOU,2020

The sign language is used widely by people who are deaf-dumb these are used as a medium for communication. A sign language is nothing but composed of various gestures formed by different shapes of hand, its movements, orientations as well as the facial expressions. There are around 466 million people worldwide with hearing loss and 34 million of these are children. `Deaf' people have very little or no hearing ability. They use sign language for communication. People use different sign languages in different parts of the world. Compared to spoken languages they are very less in number. In existing system, lack of datasets along with variance in sign language with locality has resulted in restrained efforts in finger gesture detection. Existing project aims at taking the basic step in bridging the communication gap between normal people and deaf and dumb people using Indian sign language. Effective extension of this project to words and common expressions may not only make the deaf and dumb people communicate faster and easier with outer world, but also provide a boost in Developing autonomous systems for understanding and aiding them. The Indian Sign Language lags behind its American Counterpart as the research in this field is hampered by the lack of standard datasets. In addition to the intrinsic challenges of human motion analysis (such as variations in the participants' appearances, the characteristics of the human silhouette, and the execution of the repetition of operations, the presence of obstructions, etc.) A signer's glosses can be inferred from video recordings using a process known as sign language recognition (SLR). Despite the fact that there is a lot of labour, There is a severe paucity of comprehensive experimental research in the subject of SLR. Additionally, most articles don't release their code or present findings from all available datasets.

2.1.5 TITLE: TRANSFERRING CROSS-DOMAIN KNOWLEDGE FOR VIDEO SIGN LANGUAGE RECOGNITION

AUTHOR: DONGXU LI,2020

As a fundamental sign language interpretation task, word-level sign language recognition (WSLR) aims to help deaf people communicate. However, WSLR is very difficult because it requires quick body movements, facial expressions, and complex, fine-grained hand gestures. Isolated Sign Words Web News Sign Words Localizer has been demonstrated recently using deep learning approaches. Our model learns domain-invariant characteristics to transfer knowledge from web news signs to WSLR models. Our model recognises the example frames in the figure as the signature that best captures the gesture on the WSLR job, their advantages. Although the largest existing datasets have a limited number of instances, e.g., on average 10 to 50 instances per word, annotating WSLR datasets requires domain-specific knowledge. This is significantly less than typical video datasets on action learning and recognition, for example. The sign recognition task's inadequate training data may cause overfitting or in some other way hinder WSLR's performance.Models under realistic circumstances. On the other hand, there are many readily available news videos with subtitles available online that could be useful for WSLR. Despite the availability of sign news videos, it is quite difficult to translate this knowledge to WSLR. First, there are no annotations of temporal location or categories and just flimsy labels for the presence of signs in subtitles. Furthermore, these labels are loud. In this study, we provide a technique for transferring cross-domain knowledge from news signs to WSLR models to enhance their performance. More specifically, using a base WSLR model in a sliding window fashion, we first create a sign word localizer to extract sign words. Then, we suggest jointly coarse-aligning two domains. Employing isolated and news indicators to train a classifier. We compute and store the centroid of each class of the coarsely-aligned new words in an external memory termed prototype memory after getting the representations of the coarsely-aligned news words.

2.2 REFERENCES

[1] Mohammed Mustafa, “ A study on arabic sign language recognition for diferently abled using advanced machine learning classifers”,2020.

[2] Necati Cihan Camg¨oz, “: Sign language transformers: joint end-to-end sign language recognition and translation”,2021.
[3]Ankita Wadhawan, “Sign language recognition systems: a decade systematic literature review”,2020
[4] Nikolas Adaloglou “ a comprehensive study on sign language recognition methods”,2020.
[5] Dongxu li,“ transferring cross-domain knowledge for video sign language recognition”,2020

2.3 PROBLEM STATEMENT DEFINITION

The sign language is used widely by people who are deaf-dumb these are used as a medium for communication. A sign language is nothing but composed of various gestures formed by different shapes of hand, its movements, orientations as well as the facial expressions. There are around 466 million people worldwide with hearing loss and 34 million of these are children. `Deaf' people have very little or no hearing ability. They use sign language for communication. People use different sign languages in different parts of the world. Compared to spoken languages they are very less in number. In existing system, lack of datasets along with variance in sign language with locality has resulted in restrained efforts in finger gesture detection. Existing project aims at taking the basic step in bridging the communication gap between normal people and deaf and dumb people using Indian sign language. Effective extension of this project to words and common expressions may not only make the deaf and dumb people communicate faster and easier with outer world, but also provide a boost in Developing autonomous systems for understanding and aiding them. The Indian Sign Language lags behind its American Counterpart as the research in this field is hampered by the lack of standard datasets

3. IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS

3.2 IDEATION & BRAINSTORMING

3.3 PROPOSED SOLUTION

In computer vision-based gesture recognition, the camera is used as input, and image processing is done before recognition. The techniques utilised to recognise the processed motions after that include Neural Network approaches and the region of interest algorithm. A vision-based sign language identification system has a fundamental drawback in that the process of gathering images is sensitive to a variety of environmental factors, including camera positioning, background circumstances, and lightning sensitivity. However, it is more practical and economical than using a camera and tracker to gather information. However, neural network techniques like the Hidden Markov Model are integrated with camera data for increased accuracy.

3.4 PROBLEM SOLUTION FIT

People who are deaf-dumb frequently employ sign language as a means of communicating. A sign language is nothing more than a collection of varied hand gestures created by varying hand shapes, movements, and orientations, as well as face expressions. 34 million of the 466 million people with hearing loss in the world's population are children. People who identify as "deaf" have very little or no hearing. They communicate using sign language. Around the world, many sign languages are used by people. They are quite few in number when compared to spoken languages. In computer vision-based gesture recognition, the camera is used as input, and image processing is done before recognition. The techniques utilised to recognise the processed motions after that include Neural Network approaches and the region of interest algorithm. A vision-based sign language identification system has a fundamental drawback in that the process of gathering images is sensitive to a variety of environmental factors, including camera positioning, background circumstances, and lightning sensitivity. However, it is more practical and economical than using a camera and tracker to gather information. However, neural network techniques like the Hidden Markov Model are integrated with camera data for increased acc

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

HAND IMAGE ACQUISITION:

The hand gesture, during daily life, is a natural communication method mostly used only among people who have some difficulty in speaking or hearing. However, a human computer interaction system based on gestures has various application scenarios. In this module, we can input the hand images from real time camera. The inbuilt camera can be connected to the system. Gesture recognition has become a hot topic for decades. Nowadays two methods are used primarily to perform gesture recognition. One is based on professional, wearable electromagnetic devices, like special gloves. The other one utilizes computer vision. The former one is mainly used in the film industry. It performs well but is costly and unusable in some environment. The latter one involves image processing. However, the performance of gesture recognition directly based on the features extracted by image processing is relatively limited. Hand image captured from web camera. The purpose of Web camera is to capture the human generated hand gesture and store its image in memory. The package called python framework is used for storing image in memory

BINARIZATION

Background subtraction is one of the major tasks in the field of computer vision and image processing whose aim is to detect changes in image sequences. Background subtraction is any technique which allows an image's foreground to be extracted for further processing (object recognition etc.). Many applications do not need to know everything about the evolution of movement in a video sequence, but only require the information of changes in the scene, because an image's regions of interest are objects (humans, cars, text etc.) in its foreground. After the stage of image preprocessing (which may include image denoising, post processing like morphology etc.) object localization is required which may make use of this technique. Detecting foreground to separate these changes taking place in the foreground of the background. It is a set of techniques that typically analyze the video sequences in real time and are recorded with a stationary camera. All detection techniques are based on modeling the background of the image i.e. set the background and detect which changes occur. Defining the background can be very difficult when it contains shapes, shadows, and moving objects. In defining the background it is assumed that the stationary objects could vary in color and intensity over time. Scenarios where these techniques apply tend to be very diverse. There can be highly variable sequences, such as images with very different lighting, interiors, exteriors, quality, and noise. In addition to processing in real time, systems need to be able to adapt to these changes. The implement the techniques to extract the foreground from background image. Using Binarization approach to assign the values to background and foreground. Foreground pixels are identified in real time environments

6.2.3 REGION OF FINGER DETECTION

Segmentation refers to the process of partitioning a digital image into multiple segments. In other words, grouping of pixels into different groups is known as Segmentation. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics The division of an image into meaningful structures, image segmentation, is often an essential step in image analysis, object representation, visualization, and many other image processing tasks. But segmentation of a satellite image into differently textured regions (groups) is a difficult problem. One does not know a priori what types of textures exist in a satellite image, how many textures there are, and what regions have certain textures. The monitoring task can be performed by unsupervised segmentation and supervised segmentation techniques. A region of interest (ROI) is a subset of an image or a dataset identified for a particular purpose. In other words, region of interest (ROI) can be defined as a portion of an image which is needed to be filtered or to be performed some other operation on.

CLASSIFICATION OF FINGER GESTURES

Artificial Neural Networks (ANN) can learn and therefore can be trained to recognize patterns, find solutions, forecast future events and classify data. CNN is well documented to be used for traffic related tasks. Neural Networks learning and behavior is dependent on the way its individual computing elements are connected and by the strengths of these connections or weights. These weights can be adjusted automatically by training the network according to a specified learning rule until it performs the desired task correctly. CNN is a supervised learning method i.e. a machine learning algorithm that uses known dataset also known as training dataset. These known parameters help CNN to make predictions. Input data along with their response values are the fundamental components of a training dataset. In order to have higher predictive power and the ability to generalize for several new datasets, the best way is to use larger training datasets. The fingers can be classified by using convolutional neural network algorithm. CNN is a common method of training artificial neural networks so as to minimize the objective function. It is a supervised learning method, and is a generalization of the delta rule. It requires a dataset of the desired output for many inputs, making up the training set. It is most useful for feed-forward networks (networks that have no feedback, or simply, that have no connections that loop).

SIGN RECOGNITION

Sign Language is a well-structured code gesture, every gesture has meaning assigned to it. Sign Language is the only means of communication for deaf people. With the advancement of science and technology many techniques have been developed not only to minimize the problem of deaf people but also to implement it in different fields. From the classification of sign features, label the signs with improved accuracy rate.

4.2 NON FUNCTIONAL REQUIREMENTS

Usability

The system shall allow the users to access the system with pc using web application. The system uses a web application as an interface. The system is user friendly which makes the system easy

Availability

The system is available 100% for the user and is used 24 hrs a day and 365 days a year. The system shall be operational 24 hours a day and 7 days a week.

Scalability

Scalability is the measure of a system's ability to increase or decrease in performance and cost in response to changes in application and system processing demands.

Security

A security requirement is a statement of needed security functionality that ensures one of many different security properties of software is being satisfied.

Performance

The information is refreshed depending upon whether some updates have occurred or not in the application. The system shall respond to the member in not less than two seconds from the time of the request submittal. The system shall be allowed to take more time when doing large processing jobs. Responses to view information shall take no longer than 5 seconds to appear on the screen.

Reliability

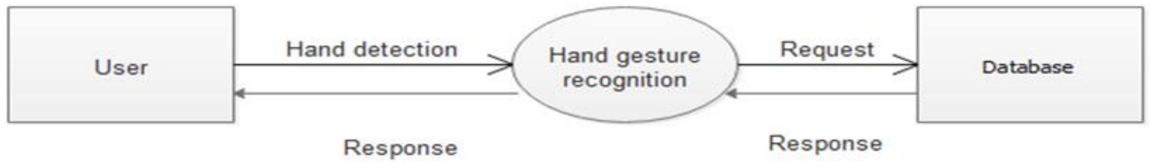
The system has to be 100% reliable due to the importance of data and the damages that can be caused by incorrect or incomplete data. The system will run 7 days a week. 24 hours a day.

5. PROJECT DESIGN

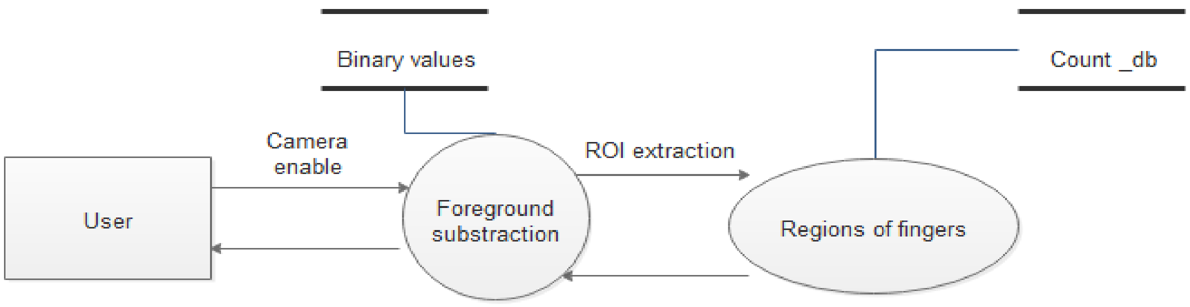
5.1 DATA FLOW DIAGRAMS

A data flow diagram is a two-dimensional diagram that explains how data is processed and transferred in a system. The graphical depiction identifies each source of data and how it interacts with other data sources to reach a common output. Individuals seeking to draft a data flow diagram must identify external inputs and outputs, determine how the inputs and outputs relate to each other, and explain with graphics how these connections relate and what they result in.

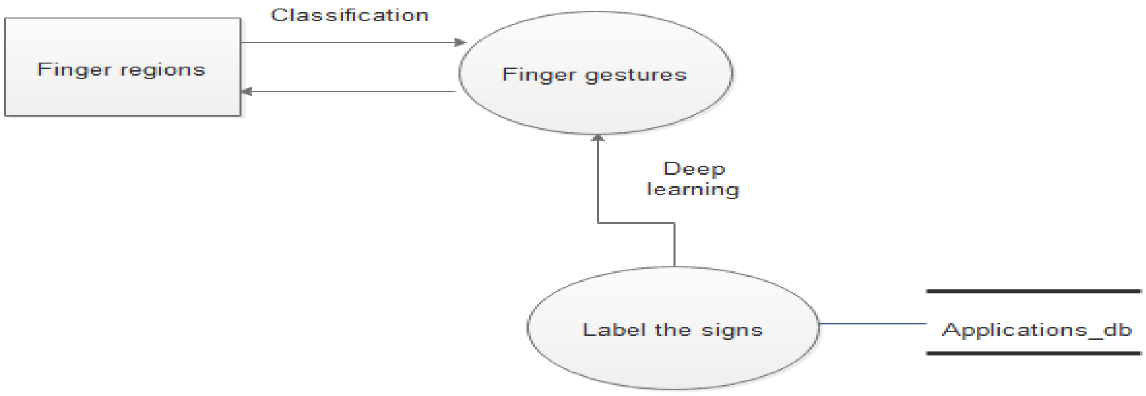
LEVEL 0:



Level 1:

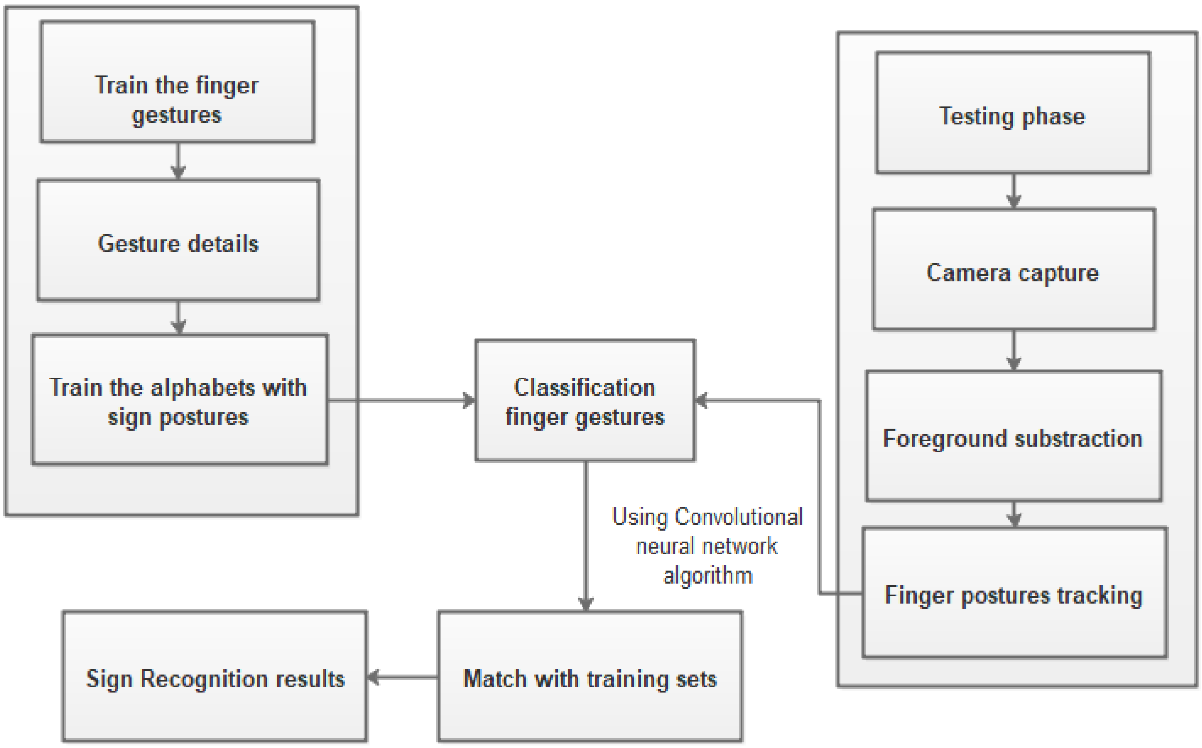


Level 2:



5.2 SOLUTION & TECHNICAL ARCHITECTURE

Software architecture involves the high level structure of software system abstraction, by using decomposition and composition, with architectural style and quality attributes. A software architecture design must conform to the major functionality and performance requirements of the system, as well as satisfy the non-functional requirements such as reliability, scalability, portability, and availability. Software architecture must describe its group of components, their connections, interactions among them and deployment configuration of all components.



5.3 USER STORIES

6. PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION

6.2 SPRINT DELIVERY SCHEDULE

6.3 REPORTS FROM JIRA

7. CODING & SOLUTION

7.1 FEATURE 1

7.2 FEATURE 2

7.3 DATABASE SCHEMA

8. TESTING

8.1 TEST CASES

A test case has components that describe input, action and an expected response, in order to determine if a feature of an application is working correctly. A test case is a set of instructions on “HOW” to validate a particular test objective/target, which when followed will tell us if the expected behavior of the system is satisfied or not.

Characteristics of a good test case:

- Accurate: Exacts the purpose.
- Economical: No unnecessary steps or words.
- Traceable: Capable of being traced to requirements.
- Repeatable: Can be used to perform the test over and over.
- Reusable: Can be reused if necessary

S.NO	**FUNCTION**	**DESCRIPTION**	**EXPECTED OUTPUT**	**ACTUAL OUTPUT**	**STATUS**
1	Framework construction	Generate the GUI for admin and user	Individual page for admin and user	Individual page for admin and user	Success
2	Read the comments	Comments Analysis	Comments in text format	Comments in text format	Success
3	Classification	Classify the Datasets	Finger Gestures	Finger Gestures	Success
4	Rules implementation	Block the comments and friends	Block the users	Block the users	Success

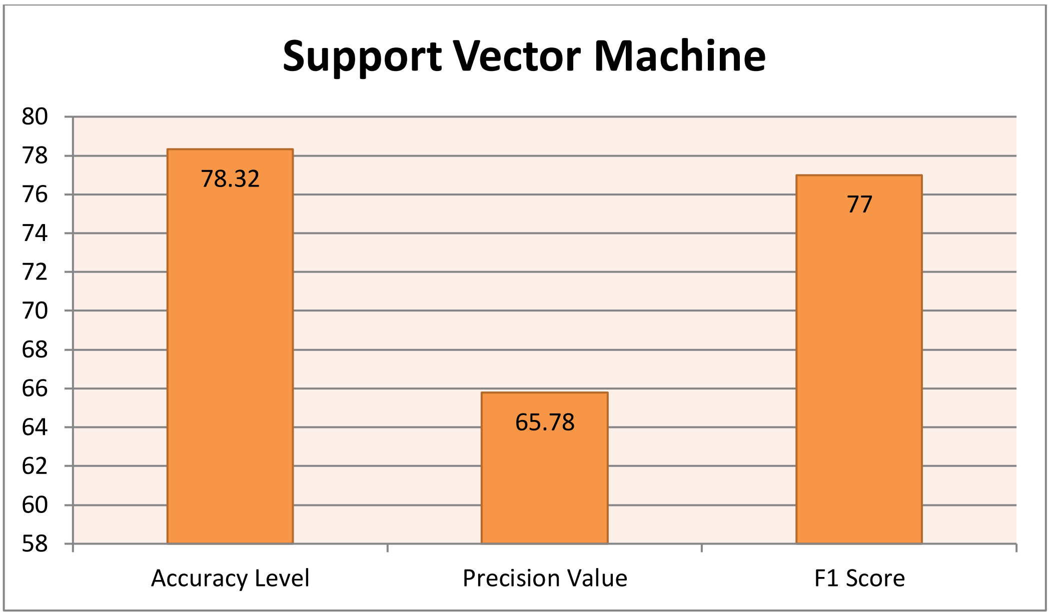
8.2 USER ACCEPTANCE TESTING

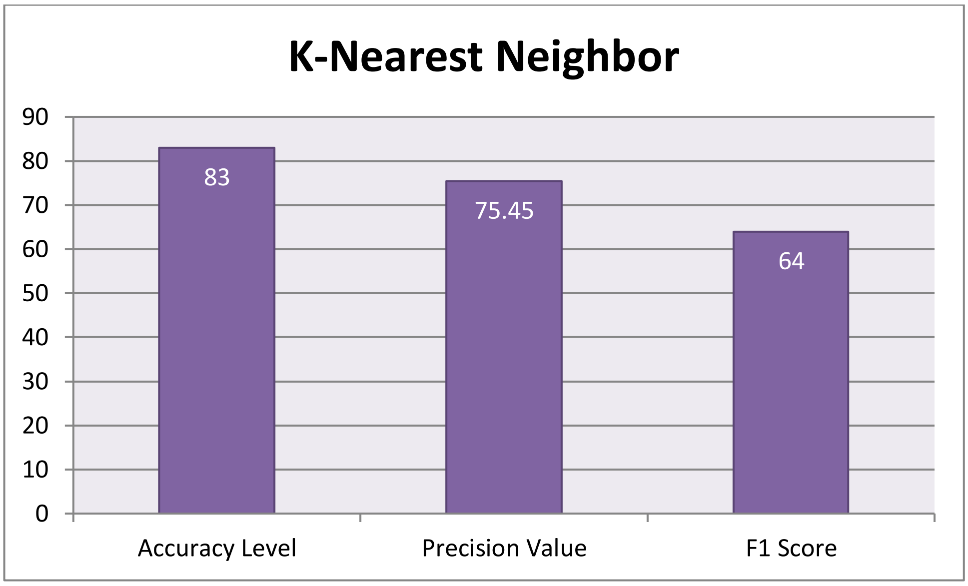
Acceptance testing can be defined in many ways, but a simple definition is the succeeds when the software functions in a manner that can be reasonable expected by the customer. After the acceptance test has been conducted, one of the two possible conditions exists. This is to fine whether the inputs are accepted by the database or other validations. For example accept only numbers in the numeric field, date format data in the date field. Also the null check for the not null fields. If any error occurs then show the error messages. The function of performance characteristics to specification and is accepted. A deviation from specification is uncovered and a deficiency list is created. User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

9. RESULTS

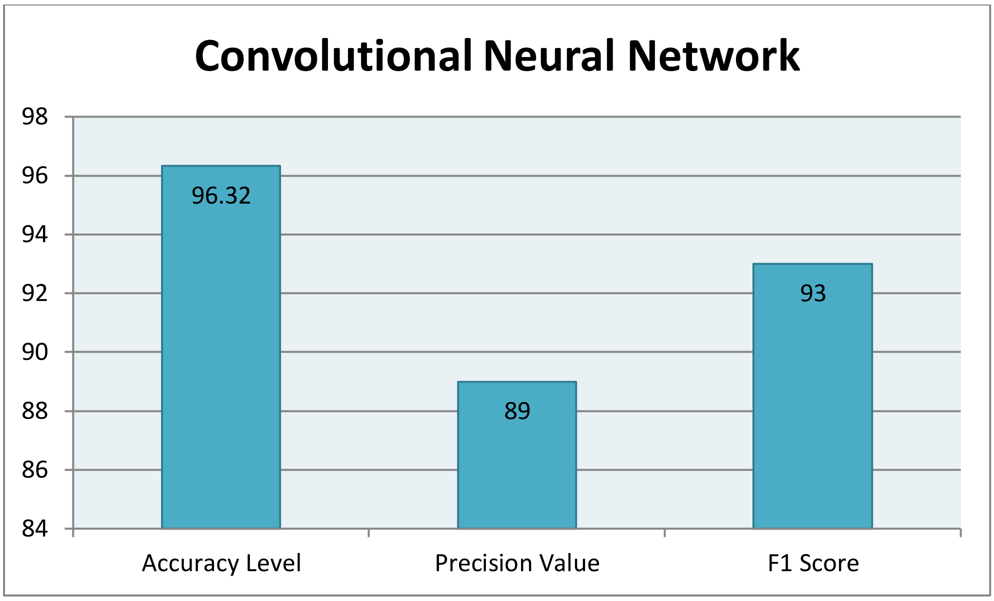
9.1 PERFORMANCE METRICS

Existing Algorithm





Proposed Algorithm



10. ADVANTAGES & DISADVANTAGES

DISADVANTAGES

- Need hardware control to detect the hands
- Hand segmentation become complex of various backgrounds
- Segmentation accuracy is less in hand tracking

ADVANTAGES

- Segmentation accuracy is high
- Easy to detect the finger postures
- Track fingers and sign recognition with less computational steps
- No need for additional hardware system

11. CONCLUSION

The ability to look, listen, talk, and respond appropriately to events is one of the most valuable gifts a human being can have. However, some unfortunate people are denied this opportunity. People get to know one another through sharing their ideas, thoughts, and experiences with others around them. There are several ways to accomplish this, the best of which is the gift of "Speech." Everyone can very persuasively transfer their thoughts and comprehend each other through speech. Our initiative intends to close the gap by including a low-cost computer into the communication chain, allowing sign language to be captured, recognised, and translated into speech for the benefit of blind individuals. An image processing technique is employed in this paper to recognise the handmade movements. This application is used to present a modern integrated planned system for hear impaired people. The camera-based zone of interest can aid in the user's data collection. Each action will be significant in its own right.

12. FUTURE SCOPE

Despite it having average accuracy, our system is still well-matched with the existing systems, given that it can perform recognition at the given accuracy with larger vocabularies and without an aid such as gloves or hand markings. In future, we can extend the framework to implement various deep learning algorithms to recognize the signs and implement in real time applications.

13. APPENDIX

SOURCE CODE

```
import csv

import copy

import cv2 as cv

import mediapipe as mp

from model import KeyPointClassifier

from app_files import calc_landmark_list, draw_info_text, draw_landmarks, get_args,
pre_process_landmark



def main():
    args = get_args()

    cap_device = args.device
    cap_width = args.width
    cap_height = args.height

    use_static_image_mode = args.use_static_image_mode
    min_detection_confidence = args.min_detection_confidence
    min_tracking_confidence = args.min_tracking_confidence

    cap = cv.VideoCapture(cap_device)
    cap.set(cv.CAP_PROP_FRAME_WIDTH, cap_width)
    cap.set(cv.CAP_PROP_FRAME_HEIGHT, cap_height)

    mp_hands = mp.solutions.hands
    hands = mp_hands.Hands(
        static_image_mode=use_static_image_mode,
        max_num_hands=1,
        min_detection_confidence=min_detection_confidence,
```

```
        min_tracking_confidence=min_tracking_confidence,
    )

    keypoint_classifier = KeyPointClassifier()

    with open('model/keypoint_classifier/keypoint_classifier_label.csv', encoding='utf-8-sig')
as f:
        keypoint_classifier_labels = csv.reader(f)
        keypoint_classifier_labels = [
            row[0] for row in keypoint_classifier_labels
        ]

    while True:
        key = cv.waitKey(10)
        if key == 27:  # ESC
            break

        ret, image = cap.read()
        if not ret:
            break
        image = cv.flip(image, 1)
        debug_image = copy.deepcopy(image)
        # print(debug_image.shape)
        # cv.imshow("debug_image",debug_image)
        image = cv.cvtColor(image, cv.COLOR_BGR2RGB)

        image.flags.writeable = False
        results = hands.process(image)
        image.flags.writeable = True

        if results.multi_hand_landmarks is not None:
```

```
        for hand_landmarks, handedness in zip(results.multi_hand_landmarks,
results.multi_handedness):
            landmark_list = calc_landmark_list(debug_image, hand_landmarks)



            #print(hand_landmarks)
            pre_processed_landmark_list = pre_process_landmark(landmark_list)

            hand_sign_id = keypoint_classifier(pre_processed_landmark_list)

            debug_image = draw_landmarks(debug_image, landmark_list)

            debug_image = draw_info_text(
                debug_image,
                handedness,
                keypoint_classifier_labels[hand_sign_id])

        cv.imshow('Hand Gesture Recognition', debug_image)

    cap.release()
    cv.destroyAllWindows()



if __name__ == '__main__':
    main()

import numpy as np
import tensorflow as tf



class KeyPointClassifier(object):
```

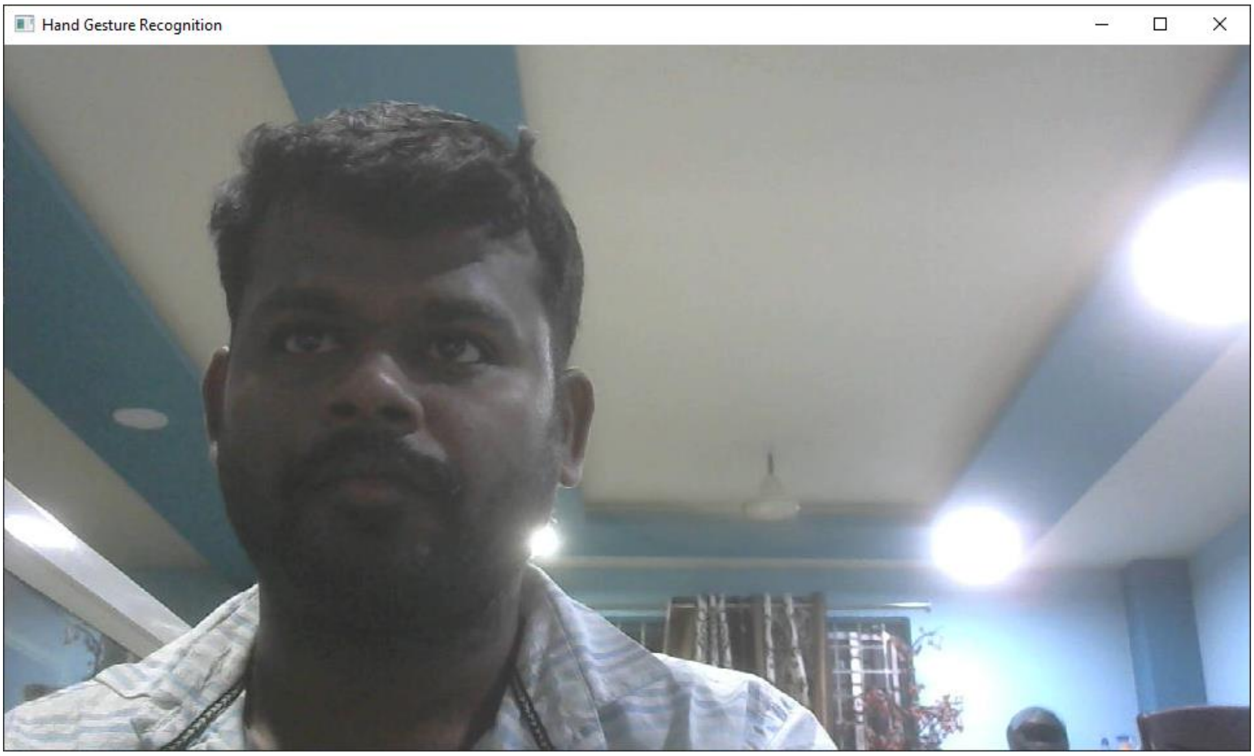
```
def __init__(
    self,
    model_path='model/keypoint_classifier/keypoint_classifier.tflite',
    num_threads=1,
):
    self.interpreter = tf.lite.Interpreter(model_path=model_path,
                                           num_threads=num_threads)

    self.interpreter.allocate_tensors()
    self.input_details = self.interpreter.get_input_details()
    self.output_details = self.interpreter.get_output_details()

def __call__(
    self,
    landmark_list,
):
    input_details_tensor_index = self.input_details[0]['index']
    self.interpreter.set_tensor(
        input_details_tensor_index,
        np.array([landmark_list], dtype=np.float32))
    self.interpreter.invoke()

    output_details_tensor_index = self.output_details[0]['index']

    result = self.interpreter.get_tensor(output_details_tensor_index)

    result_index = np.argmax(np.squeeze(result))

    return result_index
```

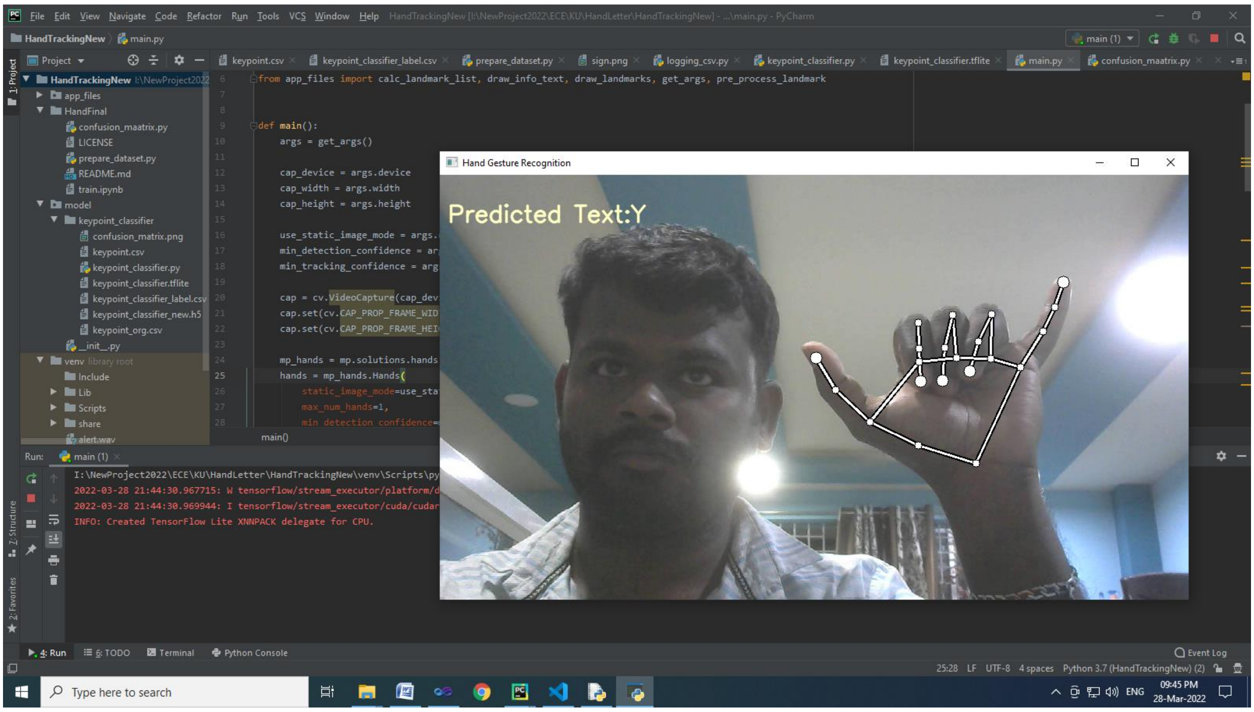
```
import copy
import cv2 as cv
import mediapipe as mp
from app_files import calc_landmark_list, draw_landmarks, get_args,
pre_process_landmark,logging_csv


def main():
    args = get_args()

    cap_device = args.device
    cap_width = args.width
    cap_height = args.height

    use_static_image_mode = args.use_static_image_mode
    min_detection_confidence = args.min_detection_confidence
    min_tracking_confidence = args.min_tracking_confidence

    cap = cv.VideoCapture(cap_device)
    cap.set(cv.CAP_PROP_FRAME_WIDTH, cap_width)
    cap.set(cv.CAP_PROP_FRAME_HEIGHT, cap_height)

    mp_hands = mp.solutions.hands
    hands = mp_hands.Hands(
        static_image_mode=use_static_image_mode,
        max_num_hands=1,
        min_detection_confidence=min_detection_confidence,
        min_tracking_confidence=min_tracking_confidence,
    )

    mode = 1
    number = -1
```

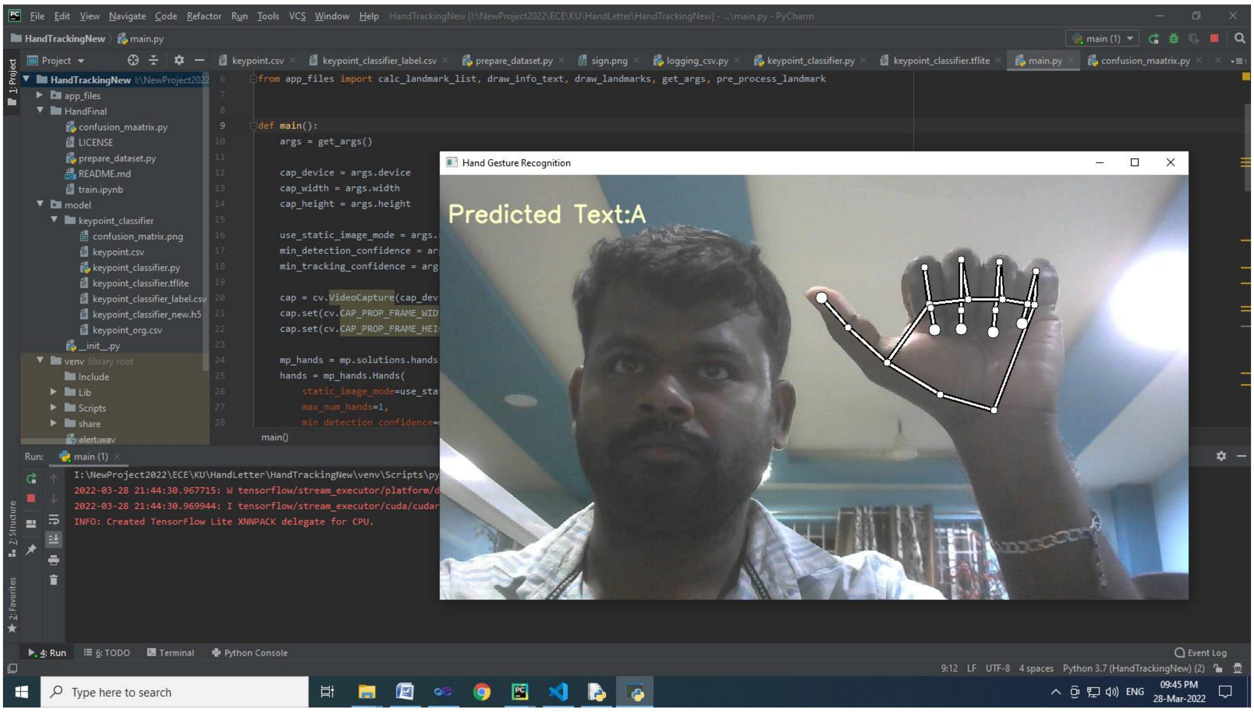
```
while True:
    key = cv.waitKey(10)
    if key == 27:  # ESC
        break

    if 48 <= key <= 57:  # 0 ~ 9
        number = key - 48

    ret, image = cap.read()

    if not ret:
        break
    image = cv.flip(image, 1)
    debug_image = copy.deepcopy(image)
    image = cv.cvtColor(image, cv.COLOR_BGR2RGB)
    image.flags.writeable = False
    results = hands.process(image)
    image.flags.writeable = True
    # cv.imshow("results",results)
    # print(type(results))
    # print(results)

    if results.multi_hand_landmarks is not None:
        for hand_landmarks in results.multi_hand_landmarks :
            landmark_list = calc_landmark_list(debug_image, hand_landmarks)
            pre_processed_landmark_list = pre_process_landmark(landmark_list)
            logging_csv(number, mode, pre_processed_landmark_list)
            # logging_csv(number, mode, landmark_list)
            debug_image = draw_landmarks(debug_image, landmark_list)
            info_text="Press key 0-9"
```

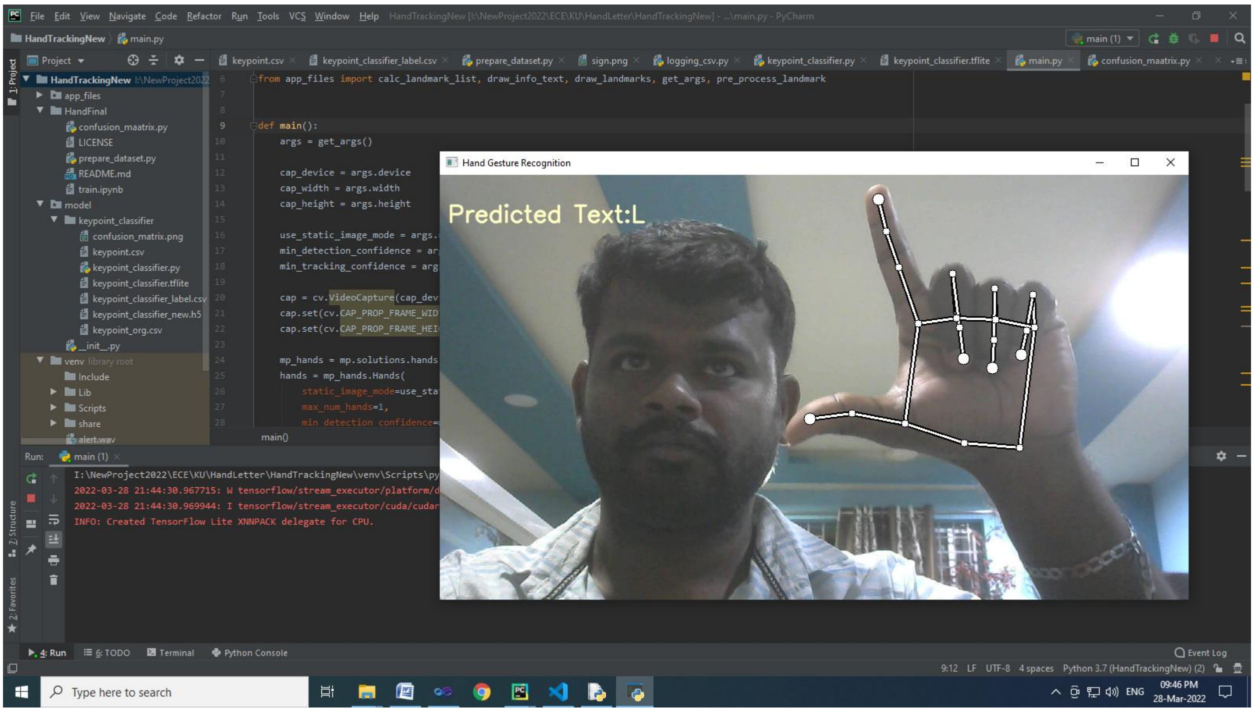
```
            cv.putText(debug_image, info_text, (10, 60), cv.FONT_HERSHEY_SIMPLEX,
1.0, (196, 161, 33), 1, cv.LINE_AA)
            cv.imshow('Dataset Preparation', debug_image)

    cap.release()
    cv.destroyAllWindows()

if __name__ == '__main__':
    main()
```

7.2 SCREENSHOTS






Hand Gesture Recognition
Predicted Text:A


Hand Gesture Recognition
Predicted Text:L

GITHUB & PROJECT DEMO LINK

GITHUB LINK:

https://github.com/IBM-EPBL/IBM-Project-43849-1660719976

DEMO LINK:

https://drive.google.com/file/d/1_VYxkmMmQ4D6R6h7xwqd2BW0w4j0OZfE/view?usp=drivesdk