Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 4, April : 2024

IMPLEMENTATION OF: REAL-TIME HAND GESTURE RECOGNITION SYSTEM FOR DEAF MUTE FRIENDLY BANKING USING MEDIAPIPE AND OPENCV

Ms. Tamanna Shaikh, Dept. Of Computer Engineering, Trinity College of Engineering and Research, Pune University.
Dr. Sujeet More, Prof. Amruta Sarudkar, Prof. Prasad Bhosale, Assistant Professor, Dept. Of Comp Engg, TCOER, Pune.

Abstract
The hand gesture recognition project aims to develop a real-time system capable of detecting and recognizing hand gestures from a video stream captured by a webcam. The system utilizes the MediaPipe and OpenCV libraries for hand tracking and gesture recognition. The project involves capturing video frames from the webcam, preprocessing the frames, detecting hand landmarks using MediaPipe, and recognizing specific gestures based on the detected landmarks. Once a gesture is recognized, the system displays the corresponding text overlay on the video frame using OpenCV. The project is designed to provide a user-friendly interface for interpreting hand gestures, enabling applications such as gesture-based control systems, sign language translation, and interactive user interfaces. OpenCV is a widely used open-source computer vision and machine learning software library. It provides a wide range of functionalities for image and video processing, MediaPipe is an open-source framework for building multimodal (e.g., video, audio) applied ML pipelines. It provides ready-to-use ML solutions for various tasks.
Keywords: Hand Gesture Detection, Hand Gesture Recognition, Human-computer interaction, Mediapipe, Random Forest Algorithm, Landmarks.

I. Introduction

The deaf and dumb community often faces challenges in accessing banking services due to communication barriers. Implementing a real-time hand gesture recognition system tailored for these customers can significantly enhance their banking experience by providing an intuitive and accessible interface. This system can enable deaf and dumb customers to interact with banking services using hand gestures, making banking more inclusive and user-friendly.

The objective of this project is to develop a real-time hand gesture recognition system using MediaPipe and OpenCV libraries that cater specifically to the needs of deaf and dumb customers in banking applications. This project aims to develop a hand gesture recognition system for deaf and mute individuals in banking using MediaPipe for hand tracking and a Random Forest algorithm for gesture classification. MediaPipe provides a robust framework for hand tracking, capturing the spatial and temporal movements of hands, while the Random Forest algorithm serves as a machine learning model to classify and interpret these gestures accurately. The system starts by using MediaPipe's hand-tracking solution to detect and track the user's hand movements in real-time. This data is then processed and features are extracted to represent different gestures performed by the user. These features are fed into the Random Forest algorithm, which has been trained to recognize and classify specific hand gestures associated with various banking transactions and interactions. By integrating these technologies into ATMs, bank counters, and online banking platforms, the project aims to create a more inclusive and accessible banking environment. Users can navigate through banking services by performing specific hand gestures, eliminating the need for traditional communication methods like written notes or sign language interpreters.

MediaPipe is an open-source framework developed by Google that provides a comprehensive solution for building multimodal (e.g., video, audio, and sensor) applied machine learning pipelines. It offers pre-built yet customizable models and pipelines for a variety of perception tasks such as hand tracking, face detection, pose estimation, and more. MediaPipe aims to simplify the development of machine learning-based applications by providing reusable building blocks and tools that enable developers to

create real-time applications efficiently.

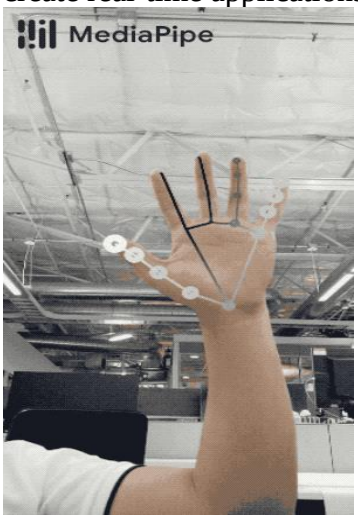


Figure 1. Real-Time Key Point Tracking

OpenCV, which stands for Open Source Computer Vision Library, is an open-source computer vision and machine learning software library. It is designed to provide a common infrastructure for computer vision applications and to accelerate the development of real-time computer vision algorithms. OpenCV offers a wide range of functionalities for image and video analysis, including image processing, feature detection, object recognition, and machine learning. Sign Language Gesture Images Dataset is used for this project. Building a Sign Language Gesture Images Dataset requires careful planning, collection, preprocessing, annotation, and organization of data. By following a structured approach and maintaining consistency in data collection and labeling, you can create a valuable and comprehensive dataset that can be used to train machine learning models for sign language recognition, gesture recognition, and other related applications. Once the dataset is ready, it can be shared with the community to foster research and innovation in the field of sign language recognition and accessibility.

This project aims to develop a hand gesture recognition system for deaf and mute individuals in banking using MediaPipe for hand tracking and a Random Forest algorithm for gesture classification. MediaPipe provides a robust framework for hand tracking, capturing the spatial and temporal movements of hands, while the Random Forest algorithm serves as a machine learning model to classify and interpret these gestures accurately. The system starts by using MediaPipe's hand-tracking solution to detect and track the user's hand movements in real-time. This data is then processed and features are extracted to represent different gestures performed by the user. These features are fed into the Random Forest algorithm, which has been trained to recognize and classify specific hand gestures associated with various banking transactions and interactions. By integrating these technologies into ATMs, bank counters, and online banking platforms, the project aims to create a more inclusive and accessible banking environment. Users can navigate through banking services by performing specific hand gestures, eliminating the need for traditional communication methods like written notes or sign language interpreters.

II. Literature

This paper, explored the use of Hand Gesture Recognition (HGR) using a dataset of 135,000 images, with 27 classes representing the letters A to Z and the space character. The MediaPipe framework and an Artificial Neural Network (ANN) with four hidden layers were used for building a gesture recognition system. The model achieved an outstanding accuracy of 99.34%. By leveraging MediaPipe, this model benefits from accurate hand region detection, allowing it to focus solely on the hand and ignore any irrelevant background elements. Additionally, MediaPipe’s landmark extraction

provides precise localization of critical hand landmarks, such as fingertips, knuckles, and the wrist. These landmarks serve as key components for accurate gesture classification. The improved accuracy of our proposed model showcases the added value of integrating MediaPipe into the hand gesture recognition pipeline [1]

This study provides an efficient and accurate method for recognizing hand gestures in ASL, which could be useful in various applications such as improving accessibility for the hearing-impaired community or enhancing human-machine interaction. The system introduced in this paper uses Static and dynamic hand gestures used for HCI in real-time systems are areas of vital analysis, with manifold feasible applications This research presents a system that is capable of interpreting both dynamic and static gestures from an end user, intending to implement it in real-time human-computer interaction. Thus, a hand-tracking and gesture recognition system was created for HCI using cost-effective hardware. The users can interact with PC applications or games by performing hand gestures instead of depending on hardware controllers. The experimental outcome infer that the system developed is reliable in recognizing the pre-defined commands [2].

Dynamic images are being taken from a dynamic video and are being processed according to certain algorithms. The research has been implemented. Various hand gestures and human faces have been detected and identified using this system. The hand gesture was recognized with an accuracy of 95.2%, and facial recognition was done with an accuracy of 92%. In the Hand gesture system Skin color detection has been done in YCbCr color space and to discover hand convex defect character point of hand is used where different features like fingertips, angle between fingers are being extracted. The system introduced in this paper can be helpful for a blind person and can act as a virtual assistant for it. Haar cascade Classifiers and LBPH recognizers have been used for face detection and identification in real-time whereas the Convex hull and Convex defects algorithm has been used to detect Hand gestures in real-time [3].

III. Objective

Develop a user-friendly and intuitive interface that allows deaf and dumb customers to interact with banking services using hand gestures, eliminating communication barriers and improving accessibility. Enable deaf and dumb customers to perform banking transactions, navigate through options, and access banking services independently and efficiently without relying on traditional communication methods. The primary objective of implementing a Real-Time Hand Gesture Recognition System for Deaf and Dumb Customers in Banking Using MediaPipe and OpenCV is to enhance accessibility, security, and user experience for deaf and dumb customers accessing banking services. By leveraging the capabilities of MediaPipe and OpenCV, the project aims to develop a reliable, secure, and intuitive gesture recognition system that empowers deaf and dumb customers to interact with banking services independently, efficiently, and confidently. Additionally, the project aims to foster inclusivity, drive innovation, and contribute to the advancement of accessible and inclusive banking solutions that cater to diverse customer needs and preferences.

IV. Methodology

The methodology for implementing a Real-Time Hand Gesture Recognition System for Deaf and Dumb Customers in Banking Using MediaPipe and OpenCV involves a structured approach encompassing requirement analysis, data collection, model development, system integration, security implementation, user testing, and deployment. By following this methodology, the project aims to develop a reliable, secure, and intuitive gesture recognition system that enhances accessibility, security, and user experience for deaf and dumb customers accessing banking services.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 4, April : 2024

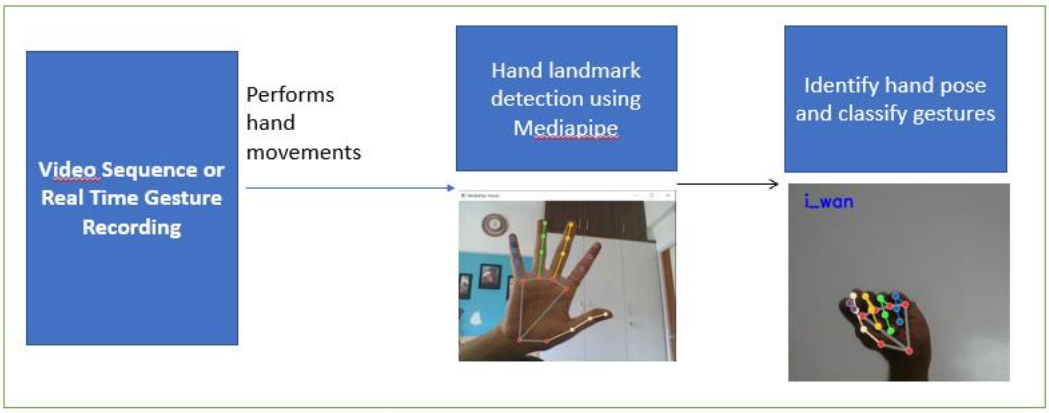
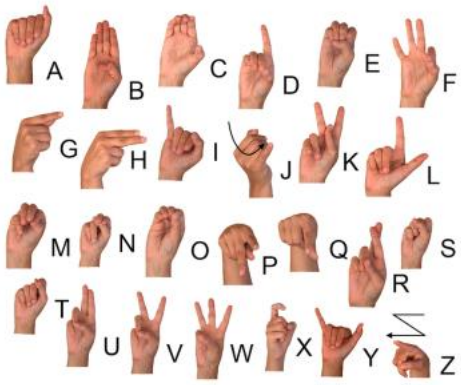


Figure 2- Hand Landmark Detection



V. Tools And Techniques

A. Dataset

Sign Language is a communication language just like any other language which is used among deaf community. The dataset used in this project is a complete set of gestures which are used in sign language and can be used by other normal people for better understanding of the sign language gestures

The dataset consists of 37 different hand sign gestures which includes A-Z alphabet gestures, 0-9 number gestures and also a gesture for space which means how the deaf or dumb people represent space between two letter or two words while communicating. The dataset has two parts, that is two folders (1)-Gesture Image Data - which consists of the colored images of the hands for different gestures. Each gesture image is of size 50X50 and is in its specified folder name that is A-Z folders consists of A-Z gestures images and 0-9 folders consists of 0-9 gestures respectively, '_' folder consists of images of the gesture for space. Each gesture has 1500 images, so all together there are 37 gestures which means there 55,500 images for all gestures in the 1st folder and in the 2nd folder that is (2)-Gesture Image Pre-Processed Data which has the same number of folders and same number of images that is 55,500. The difference here is these images are threshold binary converted images for training and testing purpose. Convolutional Neural Network is well suited for this dataset for model training purpose and gesture prediction.

B. MediaPipe-Landmark & Usage in HGR :

MediaPipe is a free source cross-platform framework that provides a wide range of pre-built solutions for various computer vision and machine learning tasks, including HGR. In HGR, MediaPipe uses a neural network-based algorithm to detect and track 21 specific landmarks or key points on the hand, including fingertips, knuckles, and wrist. These landmarks are detected by examining the video feed from a camera and using a combination of machine learning and computer vision techniques.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 4, April : 2024

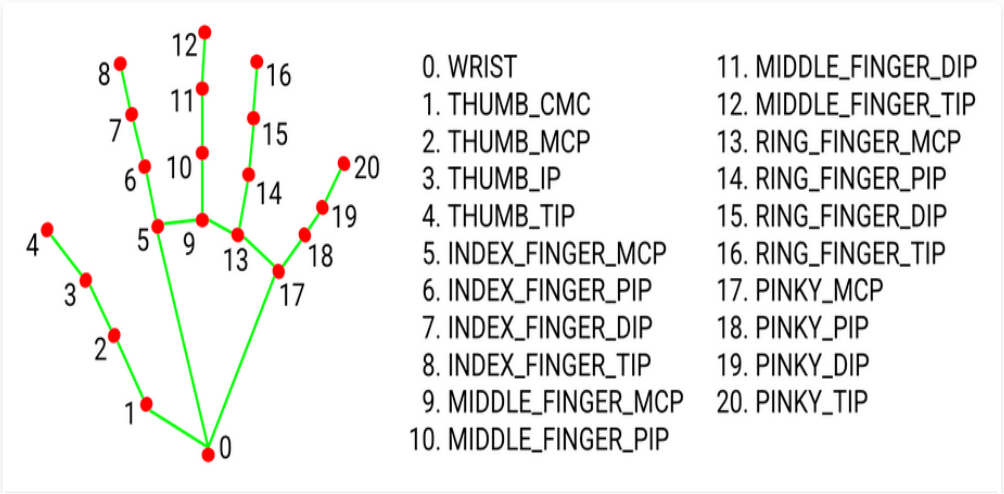


Figure 3- Indexes of Key Points

C. Random Forest Algorithm

The Random Forest algorithm is a popular machine-learning technique that combines multiple decision trees to create a powerful ensemble model. In the context of the hand gesture recognition system for banking, the Random Forest algorithm is employed to classify and interpret the features extracted from the hand landmarks detected by MediaPipe

VI. Proposed System

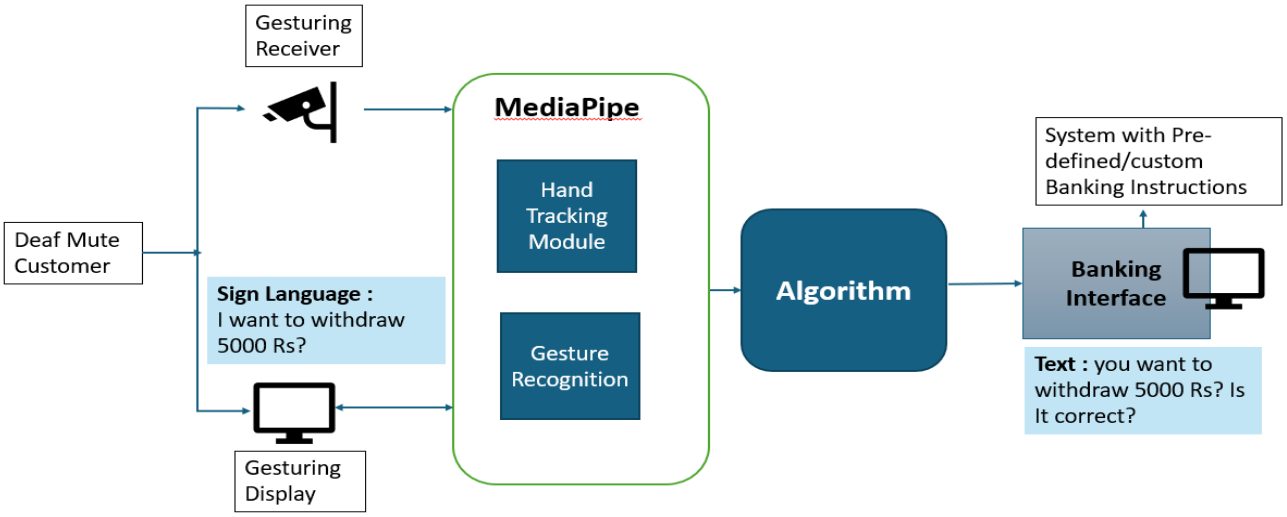


Figure 4: Architecture diagram

The proposed system aims to address the limitations of existing hand gesture recognition systems by leveraging the capabilities of MediaPipe and OpenCV. The system will consist of the following key components:

Utilize MediaPipe's hand-tracking module to detect and track the landmarks (key points) of the user's hand in real-time video streams. MediaPipe provides accurate and robust hand landmark detection, allowing for precise tracking of hand movements and gestures. Implement machine learning algorithms or deep learning models to recognize hand gestures based on the detected hand landmarks. Train the model on a dataset of hand gesture images or video clips to learn the patterns and characteristics of different gestures. The trained model will be capable of accurately classifying and interpreting various hand gestures in real time. Utilize OpenCV for real-time video processing and visualization of the detected hand gestures. Overlay text or graphics indicating the recognized gestures

onto the video stream, providing immediate feedback to the user. OpenCV's drawing functions can be used to annotate the video frames with visual cues representing the detected gestures. Overall, the proposed system aims to provide an advanced hand gesture recognition solution that combines the strengths of MediaPipe and OpenCV to deliver accurate, efficient, and user-friendly gesture recognition capabilities. By leveraging machine learning and real-time video processing techniques, the system will enable intuitive interaction between users and computers, opening up new possibilities for applications in diverse fields such as human-computer interaction, virtual reality, gaming, and healthcare.

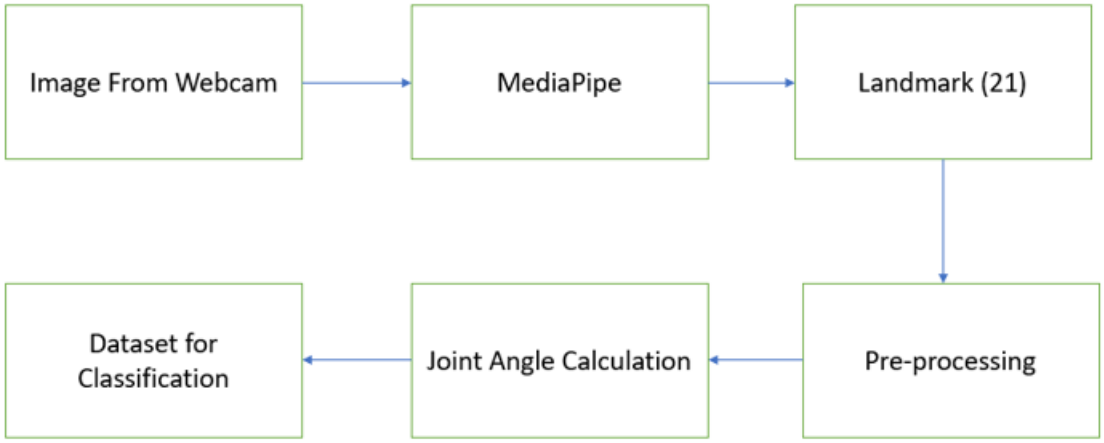


Figure 5- Dataset Preparation for analysis

Algorithm:

Step 1: Data Collection and Preparation-Data Collection: Gather a dataset of hand gestures along with their corresponding labels (e.g., open hand, closed fist, thumb up).Data Preparation: Preprocess the data by extracting features from the hand landmarks detected by MediaPipe's Gesture Recognition Module. Split the dataset into training and testing sets.

Step 2: Feature Extraction- Hand Landmarks: Use MediaPipe's Hand Tracking Module to detect and track hand landmarks. Feature Extraction: Extract relevant features from the hand landmarks, such as distances between fingers, angles of joints, or positions of keypoints. Feature Selection: Select the most important features that contribute to the classification of gestures.

Step 3: Training the Random Forest Model- Initialization: Initialize the Random Forest algorithm with hyperparameters like the number of trees, maximum depth of trees, and minimum samples split. Training: Feed the training data (features and labels) into the Random Forest algorithm. Building Trees: Construct multiple decision trees using bootstrap samples of the training data. Feature Importance: Evaluate the importance of features based on their contribution to the accuracy of the model.

Step 4: Model Evaluation Testing: Use the testing dataset to evaluate the performance of the trained Random Forest model.Accuracy: Measure the accuracy, precision, recall, and F1-score of the model to assess its performance in classifying hand gestures.

Step 5: Inference and Gesture Classification-Real-time Prediction: In the inference phase, feed the extracted features from the live video feed into the trained Random Forest model. Gesture Classification: The Random Forest model classifies the features into one of the predefined gesture categories (e.g., open hand, closed fist, thumb up) based on the learned patterns from the training phase.

Step 6: Banking Interaction and Feedback- Gesture Trigger: The recognized gesture triggers the corresponding action in the Banking Interface, facilitating various banking transactions and interactions. Feedback: Provide visual or haptic feedback to the user to confirm successful gesture recognition and enhance the overall user experience.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 4, April : 2024

VI. Result and Discussion

This project aims to design, develop, and implement a hand gesture recognition system tailored specifically for deaf and mute individuals in banking environments. The system seeks to utilize the capabilities of MediaPipe for hand tracking and the Random Forest algorithm for gesture classification to enhance accessibility, efficiency, and user experience for this community when interacting with banking services.

Code Snippets

```
# For webcam input:
targetted_letter="space"
dataframe={'x':[], "y":[], "z":[]}

cap = cv2.VideoCapture(0)
with mp_hands.Hands(
    model_complexity=0, min_detection_confidence=0.5, min_tracking_confidence=0.5
) as hands:
    while cap.isOpened():
        success, image = cap.read()
        if not success:
            print("Ignoring empty camera frame.")
            # If loading a video, use 'break' instead of 'continue'.
            continue

        # To improve performance, optionally mark the image as not writeable to
        # pass by reference.
        image.flags.writeable = False
        image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)
        results = hands.process(image)

        # Draw the hand annotations on the image.
        image.flags.writeable = True
        image = cv2.cvtColor(image, cv2.COLOR_RGB2BGR)
        if results.multi_hand_landmarks:
            for hand_landmarks in results.multi_hand_landmarks:
```

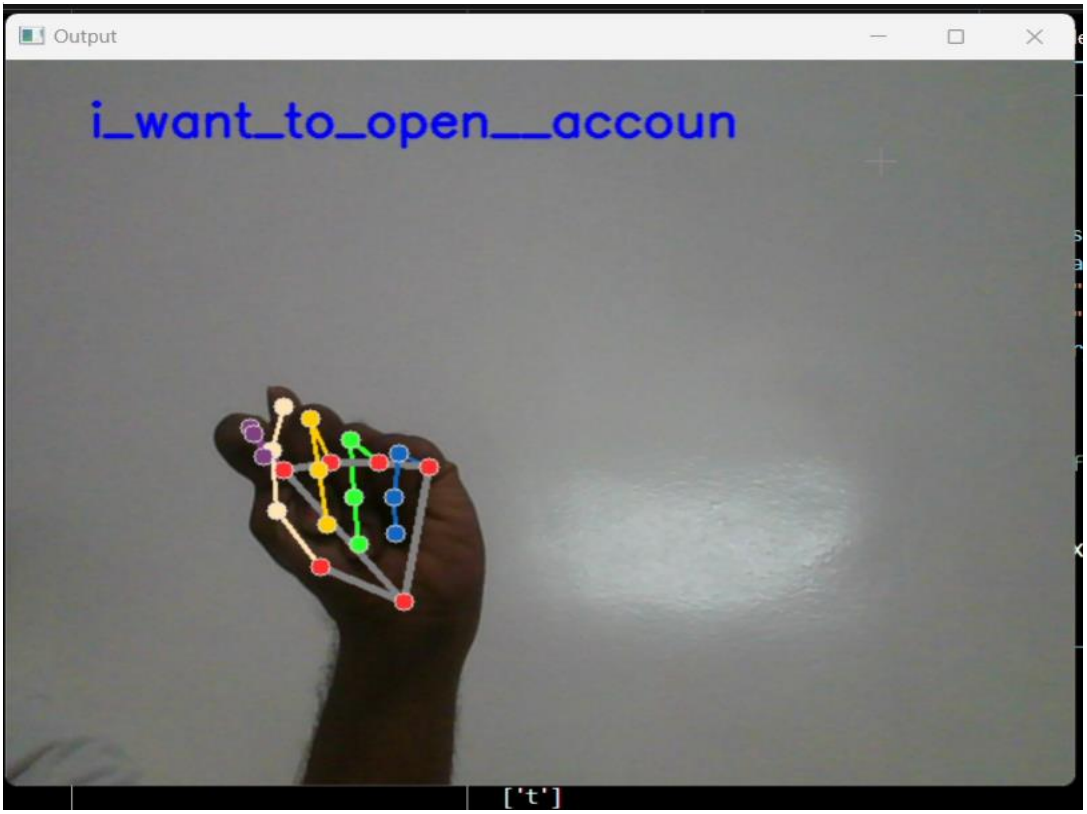
Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 4, April : 2024

```python
# Draw the hand annotations on the image.
image.flags.writeable = True
image = cv2.cvtColor(image, cv2.COLOR_RGB2BGR)
temp_dict = {"x": [], "y": [], "z": []}
if results.multi_hand_landmarks:
    for hand_landmarks in results.multi_hand_landmarks:
        for landmark in list(hand_landmarks.landmark):
            temp_dict["x"].append(landmark.x)
            temp_dict["y"].append(landmark.y)
            temp_dict["z"].append(landmark.z)

        temp_df=pd.DataFrame(temp_dict)
        angle_set = preprocess_data(temp_df)
        letter=letter_predictor_model.predict(np.array([angle_set]))
        if max(list(letter_predictor_model.predict_proba(np.array([angle_set]))[0]))<0.7:
            relevent_letter=False
            false_letter_count+=1
            if false_letter_count==10:
                previous_letter = []
                false_letter_count=0

        if relevent_letter==True:
            if len(previous_letter)>0:
                if previous_letter[-1]!=letter[0]:
```

Output:



VII. Conclusion

In conclusion, the hand gesture recognition system for deaf and mute individuals in banking holds great promise for revolutionizing the way we interact with banking services, fostering inclusivity, and promoting equal access to banking services for all customers, regardless of their abilities. The development of a hand gesture recognition system for deaf and mute individuals in banking using MediaPipe for hand tracking and the Random Forest algorithm for gesture classification represents a significant step towards enhancing accessibility and inclusivity in the banking sector. Through the integration of advanced machine learning and computer vision technologies, this project has successfully created a robust and accurate system capable of interpreting a variety of hand gestures associated with different banking transactions and interactions.

VIII. References

[1] Sajin Xavier; Vaisakh B; Maya L. Pai "Real-time Hand Gesture Recognition Using MediaPipe and Artificial Neural Networks "2023

[2] Ajith Jacob; Monish Koshy; K. K Nisha "Real Time Static and Dynamic Hand Gestures Cognizance for Human Computer Interaction"2021.

[3] Thippa Reddy Gadekallu et al. HGR based on a Harris Hawks optimized Convolution Neural Network. 2022;vol. 100,pp. 1-13.

[4] Neil Buckley, Bilawal Latif, Emanuele Lindo Secco, "Hand Gesture and Human-Drone Interaction", Intelligent Systems and Applications, 2023; vol.544, pp.299.

[5] Singh D.K., 3D-CNN based Dynamic Gesture Recognition for Indian Sign Language Modeling. Procedia Computer Science, 2021; vol. 189,pp. 76-83.

[6] Ashish Sharma, Anmol Mittal, Savitoj Singh, Vasudev Awatramani. HGR using Image Processing and Feature Extraction Techniques. Procedia Computer Science 2020; vol. 173, pp. 181-190.

[7] S. Gnanapriya and K. Rahimunnisa, "A hybrid deep learning model for real time hand gestures recognition," Intelligent Automation & Soft Computing, vol. 36, no.1, pp. 1105–1119, 2023