Proceedings of the 3rd International Conference on Inventive Computing and Informatics (ICICI-2025)
DVD Part Number: CFP25L34-DVD; ISBN: 979-8-3315-3829-3

Enhancing Security With Face Recognition and Hand Gesture

Dr M Sobhana
Department of Computer Science and Engineering
Siddhartha Academy of Higher education
Vijayawada, India
sobhana@vrsiddhartha.ac.in

Mudunuri V CH S Raju
Department of Computer Science and Engineering
Siddhartha Academy of Higher education
Vijayawada, India
satyaramaraju1234@gmail.com

Bommaganti S V Vikas
Department of Computer Science and Engineering
Siddhartha Academy of Higher education
Vijayawada, India
vikas2004bommaganti@gmail.com

Kalle Bhavya Vani
Department of Computer Science and Engineering
Siddhartha Academy of Higher education
Vijayawada, India
bhavyavanikalle@gmail.com

***Abstract*—Traditional authentication techniques like pin and password are widely used but due to its limitations and vulnerabilities, we introduced the biometric authentication techniques like face recognition, fingerprints, iris. But a single layer of authentication system can be prone to data breach. To overcome this, we proposed a dual authentication system using two types of biometrics, which are contact less and non confidential. The combination of face recognition and hand gesture can be used as a dual authentication system. The FaceNet is used for the face recognition and Mediapipe for the hand gesture recognition. Even if the face data is compromised, the hand gesture acts as an additional layer of security. By leveraging the distinctiveness of user defined gestures for various angles, this approach significantly enhances the security and makes spoofing difficult.The proposed model has achieved an accuracy of 90 percentage in recognizing the face and hand gesture.The adaptability of this solution makes it future-ready, aligning with the evolving demands of the modern workplace while ensuring robust protection against unauthorized access. Whenever the system cannot verify either face or hand gesture, the authorization will not be provided. The face embeddings generated by FaceNet of the registered faces and coordinates of the landmark of the hand gesture ensured the privacy of the user information. The proposed system works in a multi threaded environment making it suitable for real time implementation where speed and authentication is necessary. It ensures a reliable verification process.**

***Index Terms*—Face Recognition, Face Detection, Hand Gesture Recognition, Multi-thread, FaceNet, MTCNN, MediaPipe.**

I. INTRODUCTION

In the current technological world, security and authentication plays a crucial role which is evolved so faster. Advances in computer vision has changed how the users authenticate with digital devices, transitioning from the traditional pin and password methods to biometric authentication systems like face recognition, fingerprints, which offers enhanced security and ease. The requirement of hygiene in the authentication systems becomes necessary after the pandemic situation. Among the biometric authentication systems, the face recognition becomes popular due to its non intrusive nature and faster deployment. However, depending completely on one biometric authentication is not suggestable because of various vulnerabilities like spoofing, identity theft.

To overcome these challenges, this paper proposes a dual-authentication system which combines the both face recognition and hand gesture recognition to authenticate an individual as an additional security layer and maintains the logging data separately. The proposed system uses the deep learning and computer vision techniques to identify the user and authenticate them by simultaneously recognizing the user's face and pre registered hand gesture using multi-threaded environment. This dual authentication system helps to reduce the chance of spoofing and unauthorised access to confidential information.

The face recognition involves two stages. In stage one, the face from the input video stream is identified with the help of MTCNN (Multi Task cascaded Convolutional Neural Network)[1] over any other algorithms and the second stage involves creating the 128- dimensional face embeddings for the face identified in the detection stage, by using the popularand effiecient performing algorithm FaceNet[2]. And at the end, the face classified out of the faces available in the database. For Hand gesture recognition, the system uses the MediaPipe[3] for real time hand key point tracking and a custom trained neural network classifier to predict the hand gesture from the hand key-point data.

To implement in a real time scenarios like offices where authorised personnel only allowed, the system is implemented using a multithreaded architecture, where video capturing, face detection, face recognition and hand gesture detection are handled independent threads. Because the multithreaded environment makes it suitable for the fast response and real world applications such as secure logins, door access control and smart home systems. The system has been evaluated on various parameters like accuracy in face recognition, hand gesture recognition and maintaining the better performance during real time execution.

II. Literature Survey

In recent years, the usage of authentication systems developed so rapidly with the advancements in computer vision, deep learning. As the requirement of biometric authentication systems focused on face recognition, several research has done to use the face as a biometric. Maheen Zulfiqar et al. [4] presented a convolutional neural network based face recognition which actively detects the multiple faces from the input image using face detector and automated the process of extracting the facial features from the detected faces using pretrained CNN specifically designed for the face recognition. They developed a large database with over 9000 images and achieved an accuracy of 98 percent which shows that we can use face recognition as authentication system.

Mohsen Heidari et al. [5] employed a siamese network approach with transfer learning from a VGG-16 model for face recognition in this paper. Two similar CNNs process pairs of face images, utilizing a similarity criterion to determine if they belong to the same person. By using the LFW dataset the rate of accuracy is 95.62%. To advance this research, exploring alternative CNN architectures within the Siamese network, especially those capable of extracting both high and low-level features, is recommended. Furthermore, adopting the ”triplet loss” method and employing data augmentation techniques may enhance performance, particularly for small-sample datasets.

S.Fong et al.[6] proposed an authentication model to enhance the authentication system without any contact with the machine by using hand gesture recognition. The proposed system uses the images that are captured through camera.Gesture image acquisition, preprocessing, feature extraction, classification, image enhancing and edge detection are been implemented to improve gesture clarity. In this system gesture variability and background noise were recognized as the major challenges. To reduce the problem with the background noise we can use the key points of the hand instead of direct images.

L. V. Da et al. [7] investigated using MediaPipe to detect hand gestures, a real-time system devised by Google to monitor hand landmarks. MediaPipe was used by the study to detect and identify hand movements by relying on pre-existing machine learning models within MediaPipe to classify the gestures with precision. Finger locations and orientations of hands were monitored by the authors for purposes of classification. This approach proved effective in detecting static and dynamic hand gestures with low computational overhead. By considering this, we can conclude that usage of hand gesture as authentication is promising.

Bashar Saadoon Mahdi et al.[8] proposed a security model for password generation using hand gesture. The model records distinctive hand movements and converts them into secure password components based on gesture dynamics like angle, position, and motion trajectory. It analyzes the password strength using machine learning algorithms. The suggested model proved security and less risk of vulnerabilities. And it can also be used to user friendly cybersecurity solutions and secure authentication systems, password generating tools.

The hand gesture recognition should work with efficiently and maintain a light weight architecture. Manuel Gil-Martn et al.[9] proposed that Mediapipe is a robust mechanism that can detect hand landmarks dynamically which can be used for gesture classification. The study explains the speed of the coordinates changes showing an accuracy of 80 percent using a 125-timestep input window. The findings of this research suggested that even using a reduced set of hand feature can use for secure gesture authentication.

Face recognition and hand gesture recognition are two seperate fields in computer vision, where both shows significant accuracy individually. Kao et al. [10] presented a novel framework of combination of face and hand gesture recognition to increase accuracy, particularly in security based scenarios. The proposed model is used in elevators, where the user shows a hand gesture from American Sign Language to indicate the desired floor number and face recognition verifies the face to give permission for that desired floor. This setup introduces the contactless model and reinforces security by introducing the dual authentication system.

Gurlove Singh et al. [1] explored face recognition, a critical element of individual authentication. The process involves two stages one is face detection and second one is recognition. The work examined the Eigenface and Fisher face methods,focusing on digital image processing. It identified accuracy limitations,particularly in frontal view detection due to limited adaptability to scale and rotation. The work suggests integrating an eye detection system to improve performance. This system holds promise in surveillance, mugshot matching, and potential applications in ATM and home security systems, with anticipated advancements in computer vision.

III. Methodology

This section shows the working principle of the dual authentication system and explains how it works. The idea behind this model is combination of ”who you are”(Face recognition) and ”what you do” (Hand gesture) creates a better authentication system without touching than depending on one biometric system. We developed this system using MTCNN, FaceNet, Mediapipe and some Python open source tools. And the system is tested in real time, handling both face recognition and hand gesture recognition in parallel.

A. System Overview

The entire system starts with capturing of the video from webcam or a camera module. Through the video each frame is analysed and detects the user’s face and hand. Based on the facial features, the model recognizes the person and then looks for a specific hand gesture for the completion of the authentication process. When both steps succeed the user gets access.

To implement this, we used the following.

- The MTCNN module is used for the face detection in each frame.

- FaceNet is used to generate the embeddings of the each detected face.
- Support Vector Machine(SVM) is used to recognize the user in databse.
- MediaPipe is used to detect and identify the landmarks in the hand.
- A custom classifier is used to recognize the hand gesture.
- OpenCV and Threading to implement the system seamlessly.

B. Data Collection

A diverse dataset of more than 30 person faces is developed. Each image is labeled according to the individual's identity, enabling supervised learning. To improve the model's generalizability, we included diverse variations such as changes in illumination, pose, and facial expressions.Later, the images are modified to partially occluded. Fig.1 shows the face dataset of a face at various angles and with partial occlusion.

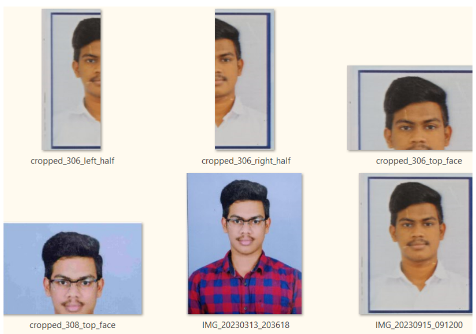

Fig. 1. Dataset for Face Recognition.

For hand gesture recognition, the dataset is developed by generating the two-dimensional coordinates, i.e, (X,Y) for each of the 21 landmarks present in the hand,stored in csv for easy access. The Fig. 2 shows the hand landmark coordinates of 14 different hand gestures each assigned to a particular user. Each user can choose a predefined hand gesture or a custom hand gesture which increases the security in authentication.

0	0	0	-0.20253	-0.10127	-0.33333	-0.29114	-0.39241	-0.46414	-0.47679	-0.57806	-0.173	-0.51055
1	0	0	-0.20307	-0.11494	-0.341	-0.29885	-0.40613	-0.4636	-0.4636	-0.59387	-0.16092	-0.52874
2	0	0	-0.17928	-0.13546	-0.28685	-0.33865	-0.32669	-0.51394	-0.38247	-0.6494	-0.08765	-0.53785
3	0	0	-0.38739	-0.14414	-0.66667	-0.4955	-0.58559	-0.81982	-0.27928	-0.81081	-0.5045	-0.75676
4	0	0	-0.4	-0.18261	-0.67826	-0.50435	-0.62609	-0.81739	-0.33913	-0.81739	-0.53043	-0.75652
5	0	0	-0.42478	-0.15929	-0.71681	-0.45133	-0.70796	-0.76106	-0.49558	-0.84071	-0.49558	-0.74336
6	0	0	-0.20161	-0.10887	-0.33065	-0.28629	-0.36694	-0.45161	-0.34677	-0.58468	-0.12903	-0.54435
7	0	0	-0.3719	-0.15702	-0.67769	-0.4876	-0.63636	-0.78512	-0.34711	-0.7686	-0.4876	-0.7438
8	0	0	-0.21569	-0.06373	-0.43627	-0.19118	-0.62745	-0.26961	-0.77941	-0.31373	-0.27451	-0.5049
9	0	0	-0.19431	-0.08531	-0.39336	-0.23223	-0.5782	-0.31754	-0.72986	-0.34597	-0.23697	-0.50237
10	0	0	-0.19608	-0.09314	-0.39706	-0.2451	-0.58824	-0.33824	-0.7402	-0.37255	-0.23529	-0.51961
11	0	0	-0.22477	-0.04128	-0.44495	-0.16972	-0.61009	-0.26147	-0.73394	-0.30275	-0.29358	-0.48624
12	0	0	-0.22222	-0.05797	-0.44928	-0.19807	-0.62802	-0.28502	-0.77295	-0.31884	-0.29469	-0.49275
13	0	0	-0.1814	-0.10698	-0.33953	-0.26047	-0.49302	-0.35814	-0.62326	-0.40465	-0.15349	-0.51628

Fig. 2. Hand Landmarks Dataset.

C. Face Recognition Pipeline

The Face recognition process starts with the detection of the face through live webcam feed. For the detection we use MTCNN model, which not only detects the face but also identify the coordinates of the precise landmarks like eyes, nose, and mouth within the frame. Number equations consecutively.

After the detection of the face, the particular region which contains the face is being cropped. resized it into 160x160 pixels and gives it to the FaceNet which generates the embedding for the face[11]. Each detected face is mapped to a 512-dimensional feature vector, which captures distinctive facial characteristics. The embeddings are further L2-normalized to maintain scale invariance and improve classification stability. This transformation ensures that similar faces are mapped closer together in the feature space, making classification more effective These embedding concentrates only on the unique features of the face, while ignoring the external features like background, light and facial expressions. The advantage of this approach is we can compare the embeddings of the detected faces without the need of the distance between the camera and user.

To recognize the face from the database, the proposed model use a SVM classifier trained on previously collected face embeddings of the users. The SVM takes the embeddings as input and predicts the probability with each face. The face with the high probability value is returned[12]. To avoid matching with random face when new face occurs, we set a threshold value of 0.9. If the face is not available in the database, it returns "Unknown".

D. Hand Gesture Recognition Pipeline

Along with the face recognition, the hand gesture recognition runs. It works using MediaPipe, a powerful real time library by Google. Primarily, it identifies the hand from the frame and detects the hand landmarks from the frame. MediaPipe keeps track of 21 keypoints[13] for each hand, including fingertips, joints, and wrist. Fig.3 shows the 21 hand landmarks depicted on the human hand. We limited the detection for one hand at a time to avoid clutter on the screen and the same gesture with different hands are identified as seperate hand gestures. That means a thumbsup gesture of left hand and right hand are not the same. Because of this, the hand gestures limit increase and brute forcing the hand gesture become difficult.

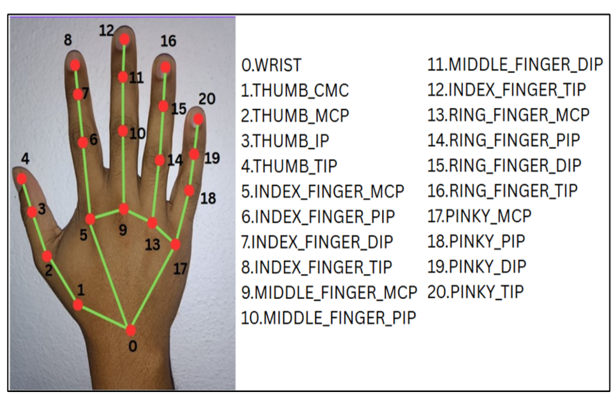

Fig. 3. Hand Landmarks Dataset.

After detecting the coordinates of all 21 landmarks, we convert them into a fixed-size vector. To avoid the issue with the distance between the camera and user, we normalize the key points relative to the wrist position of the detected hand. So the distance between the camera is not considered, only the relative distance between the points based on the wrist landmark(0 th landmark in Fig.3) is considered.

Once the normalization is completed, the landmark data is passed to the custom classifier. which we call it as "Key Point Classifier", is trained on the hand landmark dataset.where each gesture is labeled with a specific preregistered user. The classifier compares the incoming data from the webcam with the known patterns and predicts the most likely username. We stored the label data seperately, so that we can easily add new gesture easily.

E. Authentication Flow

The proposed system works in two layers as shown in the Fig.4.

- At first, the face recognition returns the name of the user using the SVM model. If the system cannot recognize the face, it immediately denies access.
- Seond, if the face is recognized, hand gesture recognition runs and returns the label of the hand gesture. The model compares both the outputs from both models.Only if both models predicted the same name, then it gives access, otherwise not.

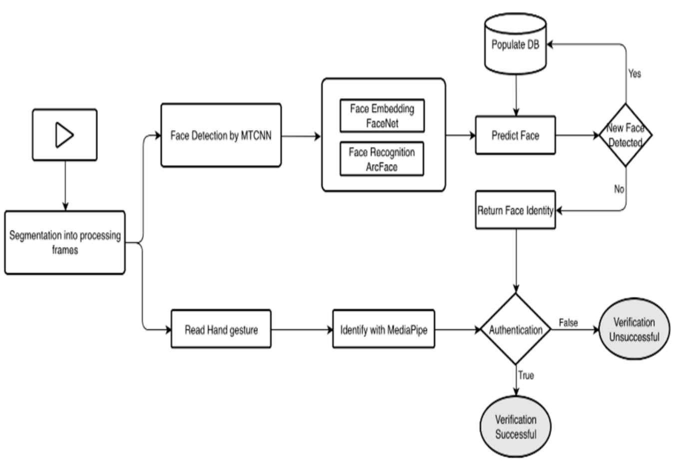

Fig. 4. Proposed Model.

IV. Results and Analysis

The proposed system is tested on using a laptop equipped with an Intel Core-i5(13th Gen), 8GB RAM and windows 11. For the video feed, we used the OpenCv library ,MTCNN and FaceNet for the face recognition and Mediapipe for the hand gesture recognition. Both face and hand recognition were executed parallely using the threading module.

To visualize the behavior of the model, we plotted training and validation loss and accuracy over 100 Epochs. The Fig 5

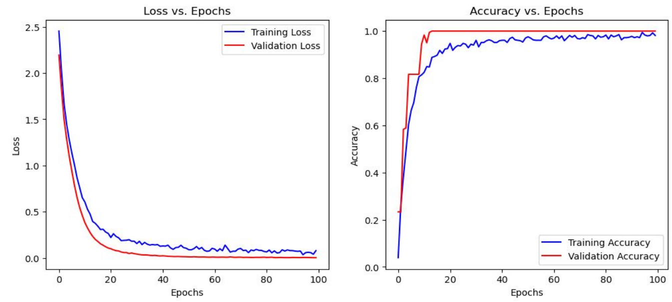

Fig. 5. Loss and Accuracy Vs Epochs

shows the loss and accuracy over epochs of the hand gesture recognition.

During the training phase, the model showed stable convergence. The validation accuracy of the model is improved consistently with a minimal overfitting. The final evaluation on the test data is achieved an accuracy approximately 96 percent.

The performance of the model is evaluate using the metrics such as Frames Per Second and execution time were monitored in real time implementation are stored in a seperate numpy files for further analysis. The Fig.6 shows the graph has been plotted showing the FPS during the execution.

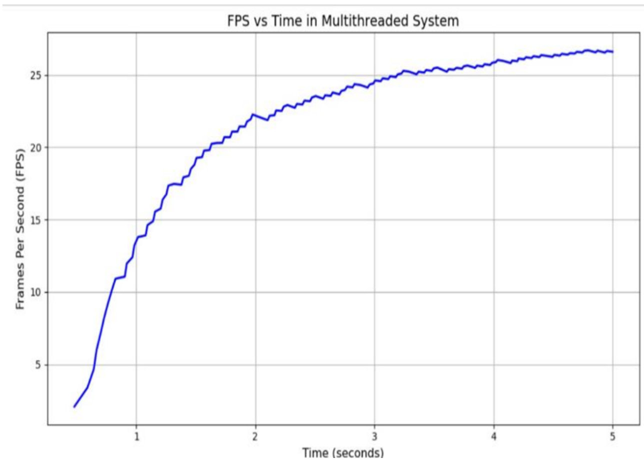

Fig. 6. FPS Vs Time

The model is tested in real time using laptop, showing good accuracy and better performance. By taking the webcam input, the model has been tested. For better understanding and evaluation, we presented the name and label of the user based on the face and hand gesture they shown. Additionally, we draw the hand skeleton[14] with the help of MediaPipe's drawing utilities as shown in Fig.7. And also presented the name of the recognized user with the help of the MTCNN for the face boundary box. This provides a clear feedback for the user to help in debugging during development.

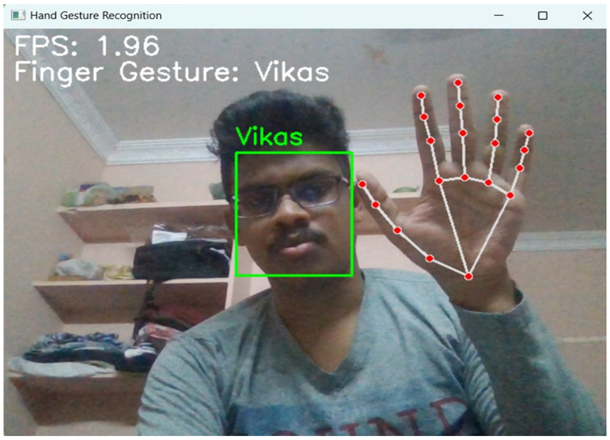

Fig. 7. Real Time Implementation

V. Future Work

The current proposed model can be developed to store user login information for further evaluation. The current model depends only on single dynamic hand gesture. In the future, we incorporate the temporal hand motion analysis technique to strengthen the authentication due to change in the hand position is also a gesture. Along with that we can add the anti-spoofing techniques like liveliness detection to ensure that the person is actually in front of the camera. Even for the authentication system, privacy is the major problem. Prioritizing privacy-preserving techniques [15] is essential to safeguard user data and mitigate potential ethical concerns. Lastly, enabling the system to continuously learn and adapt to evolving conditions and new individuals is a critical aspect of future development. By addressing these challenges and exploring these avenues, we can significantly advance the state-of-the-art in face and handgesture detection and recognition technology. Integrating this in a small scale devices like Raspberry Pi and NVIDIA Jetson can increase the usability of this model in various scenarios like smart home and surveillance applications.

VI. Acknowledgment

Since the research includes human subject data, especially the hand gestures and facial clips, for enhancement of the biometric recognition. Every individual who engaged in data collection were provided/informed consent after being briefed on the research including how their data is used, stored. These individuals were informed regarding the right to withdraw their data from research at any time without any consequences. All the procedures carried out in accordance with the informed consent and privacy protected. Although this research doesn't require formal IRB approval. This was adhered to widely accept ethical guidelines for maintaining biometric data.

References

[1] Bahmeie, A., & Shakiba, M. (2024, October). Comparing MTCNN and Viola-Jones Algorithm in Face Recognition. In 2024 19th Iranian Conference on Intelligent Systems (ICIS) (pp. 68-72). IEEE.

[2] Li, M. (2025). Research and Analysis of Facial Recognition Based on FaceNet, DeepFace, and OpenFace. In ITM Web of Conferences (Vol. 70, p. 03009). EDP Sciences.

[3] Muliawan, N. H., Irmawan, F. N., Angky, E. V., & Prabowo, A. S. (2024, November). MediaPipe's Pose-Based Human Activity Recognition With LSTM. In *2024 6th International Conference on Cybernetics and Intelligent System (ICORIS)* (pp. 1-6). IEEE.

[4] Zulfiqar, M., Syed, F., Khan, M. J., & Khurshid, K. (2019, July). Deep face recognition for biometric authentication. In 2019 international conference on electrical, communication, and computer engineering (ICECCE) (pp. 1-6). IEEE.

[5] Heidari, M., & Fouladi-Ghaleh, K. (2020, February). Using siamese networks with transfer learning for face recognition on small-samples datasets. In 2020 international conference on machine vision and image processing (MVIP) (pp. 1-4). IEEE.

[6] Fong, S., Zhuang, Y., Fister, I., & Fister, I. (2013). A biometric authentication model using hand gesture images. Biomedical engineering online, 12, 1-18.

[7] Da, L. V., Anil, C., Harish, S., & Divya, L. (2022). MediaPipe to Recognise the Hand Gestures. WSEAS TRANSACTIONS ON SIGNAL PROCESSING., 18, 134-139.

[8] Mahdi, B. S., Hadi, M. J., & Abbas, A. R. (2022). Intelligent security model for password generation and estimation using hand gesture features. Big Data and Cognitive Computing, 6(4), 116.

[9] Gil-Martín, M., Marini, M., Martín-Fernández, I., Esteban-Romero, S., & Cinque, L. (2025). Hand Gesture Recognition Using MediaPipe Landmarks and Deep Learning Networks. In In Proceedings of the 17th International Conference on Agents and Artificial Intelligence-Volume 3: ICAART (pp. 24-30).

[10] Kao, Y. W., Gu, H. Z., & Yuan, S. M. (2008, November). Integration of face and hand gesture recognition. In *2008 Third International Conference on Convergence and Hybrid Information Technology* (Vol. 1, pp. 330-335). IEEE.

[11] Ryando, C., Sigit, R., & Dewantara, B. S. B. (2024). Comparison of Machine Learning Algorithms for Face Classification Using FaceNet Embeddings. *The Indonesian Journal of Computer Science*, *13*(4).

[12] Roy, M. K., Dwibedi, P., Singh, A., Chakraborty, R. P., & Mondal, M. K. H. (2023, July). MTCNN and FACENET-Based Face Detection and Recognition Model for Attendance Monitoring. In *International Conference on Human-Centric Smart Computing* (pp. 525-539). Singapore: Springer Nature Singapore.

[13] Marais, M., Brown, D., Connan, J., & Boby, A. (2022, August). An evaluation of hand-based algorithms for sign language recognition. In 2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD) (pp. 1-6). IEEE.

[14] Veluri, R. K., Sree, S. R., Vanathi, A., Aparna, G., & Vaidya, S. P. (2022, March). Hand gesture mapping using MediaPipe algorithm. In *Proceedings of Third International Conference on Communication, Computing and Electronics Systems: ICCCES 2021* (pp. 597-614). Singapore: Springer Singapore.

[15] Figueiredo, L. S., Livshits, B., Molnar, D., & Veanes, M. (2016, May). Prepose: Privacy, security, and reliability for gesture-based programming. In 2016 IEEE Symposium on Security and Privacy (SP) (pp. 122-137). IEEE.