



A Victim Tracking System by Recognizing Signs of Violence Using Long-Short Term Memory

Hadeel Qasem Gheni¹, Noor Fadel Hussaien^{2*}, Noor Razaq²

¹ Department of Computer Science, College of Science for Women, University of Babylon, Babil 51002, Iraq

² Security Information Department, Information Technology College, University of Babylon, Babil 51002, Iraq

Corresponding Author Email: noorfadel75@gmail.com

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijse.140423>

ABSTRACT

Received: 10 June 2024

Revised: 5 August 2024

Accepted: 14 August 2024

Available online: 30 August 2024

Keywords:

LSTM, tracking violence, sign recognition, body tracking, sign for help

Artificial intelligence, especially deep learning models, is increasingly used to identify violence in many circumstances. The use of conventional cameras, mathematics-based algorithms, and deep learning models to identify hostility in real time has shown promising results. This method achieves exceptional precision without specific sensors. The study introduces a layered approach where hand gestures indicative of distress (such as requests for help) are first identified and recognized. This gesture recognition is achieved through the use of landmark points and Long Short-Term Memory (LSTM) networks to track and interpret hand motions accurately. Beyond recognizing gestures, the system advances to a second stage where it monitors and tracks violent incidents. This involves capturing images of the abused individual and triggering alarms, thus providing a comprehensive response mechanism. By integrating gesture recognition with violence tracking and employing the LSTM algorithm, the system achieves a high detection accuracy of 95%. This synergistic approach improves both the detection and tracking precision, offering a robust tool for monitoring and responding to instances of violence.

1. INTRODUCTION

Research on hand gesture recognition has greatly expanded to address various challenges in human life, especially in urgent situations such as violence, accidents, and disasters where quick and accurate actions are crucial. Effective communication tools and prompt message delivery are essential when seeking help to prevent unwanted outcomes [1].

Image processing has been used in many vital and important applications where the image gives important features through which the specific action is detected [2]. Technological advances that facilitate human-computer interaction, especially through voice and gesture interactions, have contributed significantly to addressing various human challenges. Hand gestures, being informative and consistent with everyday human habits, play a vital role in communication. Gesture recognition involves the mathematical interpretation of human movements by computing devices, which benefits individuals, especially those with physical disabilities, in communication [3].

The advancement of deep learning and computer vision has improved hand gesture recognition technologies, enabling pointing movements to be analyzed and sorted into specific regions of interest. The use of deep learning techniques, particularly in picture classification and object recognition, has hastened the development and implementation of hand gesture recognition technology. By using deep learning concepts, this technology allows computers to analyze human

hand movements and effectively convey physical or emotional reactions [4].

Hand gesture recognition models, using various devices, have revolutionized communications, allowing gestures to serve as a form of sign language to convey information. In critical situations, such as situations of violence, where rapid and accurate actions are essential, the use of appropriate communication methods and tools is crucial to ensuring prompt requests for assistance. Sign language, shown in (Figure 1), appears as a valuable means for individuals to quickly and effectively seek help when faced with danger [5].

The Sign of Help, also known as the Violence at Home Sign for Help, is a unilateral gesture that can be employed during a video chat or face-to-face interaction to notify others that someone feels endangered and requires assistance. Originally, the Sign was created as a tool to combat the rise in domestic violence cases around the world linked to self-isolation measures that were related to the COVID-19 pandemic. The Sign is executed by elevating one hand with the thumb enclosed within the palm, thereafter bending the remaining four fingers downwards, figuratively ensnaring the thumb within the grasp of the other fingers as depicted in Figure 1. The design of the hand movement was deliberately intended to be a continuous motion, rather than a static sign, in order to ensure clear visibility [5].

Segmentation methods based on traditional image processors, such as filters, sometimes give unsatisfactory results due to contrast or interference of the hand with the background [6], so landmark point overcomes these problems.

Through the calculation of the geographical location of major hand joints and the resolution of hand and finger tracking, dynamic hand motion makes use of machine learning (ML) models to estimate the hand from an image or video. We can differentiate the form, motion, and gestures of the hands with this method. It is possible to understand and follow hand gestures that communicate violence, which can be easily predicted after using such a technique [7] See (Figure 2).



Figure 1. Sign for help [5]

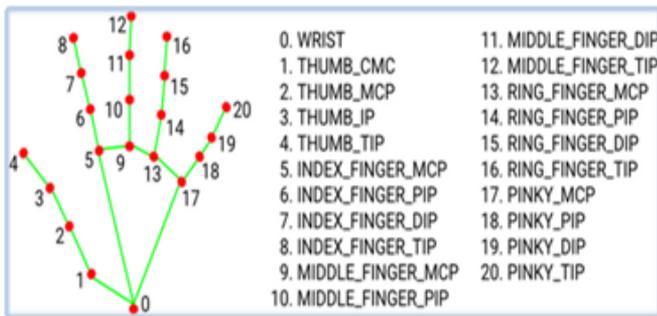


Figure 2. 21 landmarks points of hand [7]

The topic of hand gesture recognition is broad, and a lot of research has been conducted in recent years. The review indicated that most articles used a single camera (webcam or laptop) for the data acquisition process

Pose estimation estimates the 3D position and orientation of an object or person in space from a 2D picture or video feed. stance estimation can extract a person's skeletal structure and stance from an image or video for gait analysis, fall detection, and movement identification [8].

Connecting the main locations to construct the skeletal framework and tracking their movements is the tracking stage. Kalman filters and particle filters measure motion and anticipate critical point positions using probabilistic models [8, 9].

Pose estimation algorithms usually identify and track. The detection stage entails locating major joints such as the shoulders, elbows, wrists, hips, knees, and ankles (Figure 3). A deep learning model like a CNN or LSTM trained on a huge dataset of tagged photos does this [9].

The LSTM recurrent neural network (RNN) is a common deep learning architecture [10]. It captures long-term dependencies well, making it ideal for sequence prediction. Unlike other neural networks, LSTM can handle whole data sequences because to its feedback connections [11]. It is ideal for anticipating and analyzing patterns to detect and track violence. By examining sequential data and recognizing distress or emergency patterns, Long Short-Term Memory (LSTM) algorithms help detect violent hand signs like the HELP Sign and the Need Police Sign [12].

The rest of this paper is as follows: Section 2 relates to the related work. Section 3 relates to the aim of this research. Section 4 relates to the result and discussion. Section 5 relates to the conclusion.

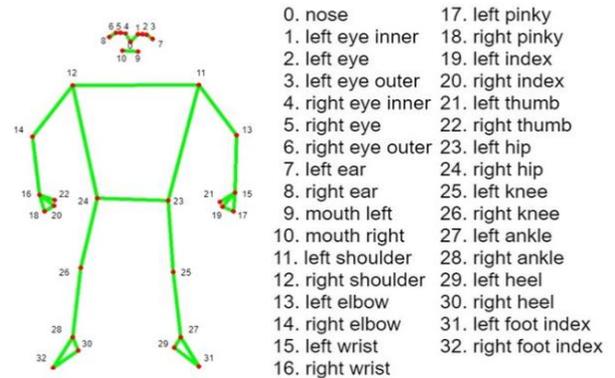


Figure 3. 33 Landmarks detected on the human body [9]

2. RELATED WORK

By integrating the open-source MediaPipe framework with the SVM learning algorithm, Taskiran et al. [13] suggested methodology streamlines real-time sign language detection in 2021. When it comes to helping the deaf communicate, the suggested paradigm is efficient, accurate, and robust. On average, it's 99% accurate.

In 2021, Mujahid et al. [14] investigated the use of hand gesture recognition to aid in communication for those with disabilities. They suggested a lightweight model that utilized YOLO (You Only Look Once) v3 and CNN for gesture detection and then measured an algorithmic accuracy of 97.68%.

Extracting and using human facial data for emotion detection and additional information for categorization was the main focus of the research. The outcomes of each processing step could be seen using (MCAD). A resolution enhancer is then applied to the picture. Identifying any humans present is the subsequent stage. After that, Haar Cascades are employed to identify faces in this area. The subsequent component processes the retrieved data to detect emotions. However, it is also feasible to combine the two blocks—emotion detection and facial point recognition—into one [15]. In this case, the coordinates of the face would no longer be used to extract emotions.

Deep learning-based real-time hand gesture recognition in complex scenarios was studied in 2019 by [4] this research aims to construct and develop a 96% accurate system for real-time hand gesture identification from video feeds using the Deep Learning CNN method.

In 2024, Zhang et al. [16] suggest a deep learning and transfer learning approach for online electromyographic hand gesture recognition. Label classifiers classify hand gesture labels; feature extractors use input signals to learn discriminant features. For online recognition, the gesture predictor employs a threshold voting method. In order to save time, the transfer learning technique moves model parameters from one pre-trained model to another. The model is tested using the NinaPro database and the Myo dataset, and it performs better than 90% of the time during typical training.

In 2023, Yu and Gu [17] provide an adaptable technique

based on deep transfer learning for precise gesture detection of a soft e-skin patch with less training data and time. The approach has the potential for extensive e-skin application in human-computer interaction, as evidenced by experimental results that ensure the stable accuracy of 95%.

3. AIM OF THE RESEARCH

This research on hand gesture recognition aims to develop a real-time system utilizing algorithms (LSTM) to enable individuals to request help urgently. Propose a hand gesture recognition system using LSTM algorithms to facilitate communication in emergencies and achieve high accuracy in real-time gesture detection. In addition, investigate the selection of good methods to detect hand gestures within various states, like illumination variation, rotation, and scaling, based on the shapes of hands by using landmark points for hands. It is worth noting that landmark point algorithms were also used to detect faces in the event of requesting help by the sign that means call the police. Here's a detailed breakdown of the problem being solved:

1. Sign language through hand gestures presents an innovative solution to overcome communication barriers in urgent situations. By recognizing specific hand movements, individuals can convey distress Signs or requests for assistance without relying on verbal communication.

2. Implementation of hand gesture recognition technology improves safety by providing a simple, fast, and safe method for individuals to request help during violent situations. Prompt message delivery, aided by the automatic capture of the victim's photo, ensures timely assistance from authorities.

Overall, the research addresses the urgent need for effective

communication tools in emergencies by leveraging hand gesture recognition technology, thereby enhancing safety and facilitating prompt assistance.

3.1 Implementation method

In this research, the general steps shown in Figure 4 illustrate the mechanism of work. The first stage involved creating a data set using landmark points for the hand and face, storing the extracted values for each point, preparing them for the LSTM application, dividing the data into training and testing, and then calculating the accuracy. This is the final step of this research.

Do Now will explore the detailed operational mechanisms of the system in real-time through the presentation of an illustrative diagram depicting the system's implementation as illustrated in (Figure 5).

The human hand was discovered using the landmark point, and once we identify the meaning of the hand sign using the LSTM, we will take a snapshot of the person exposed to violence, in addition to the possibility of tracking the person's skeletal structure. The points below reinforce the importance of using LSTM.

- Data Representation: LSTM models are trained on datasets containing examples of violent hand signs, including variations in hand movements, speed, and intensity, to learn distinguishing features.

- Temporal Analysis: LSTM's memory cells enable it to capture temporal dependencies in sequences of hand movements, allowing it to recognize characteristic patterns associated with violent signs over time. The model learns to differentiate between benign gestures and those indicative of distress or emergencies, such as sudden, forceful hand movements or repeated gestures for help.

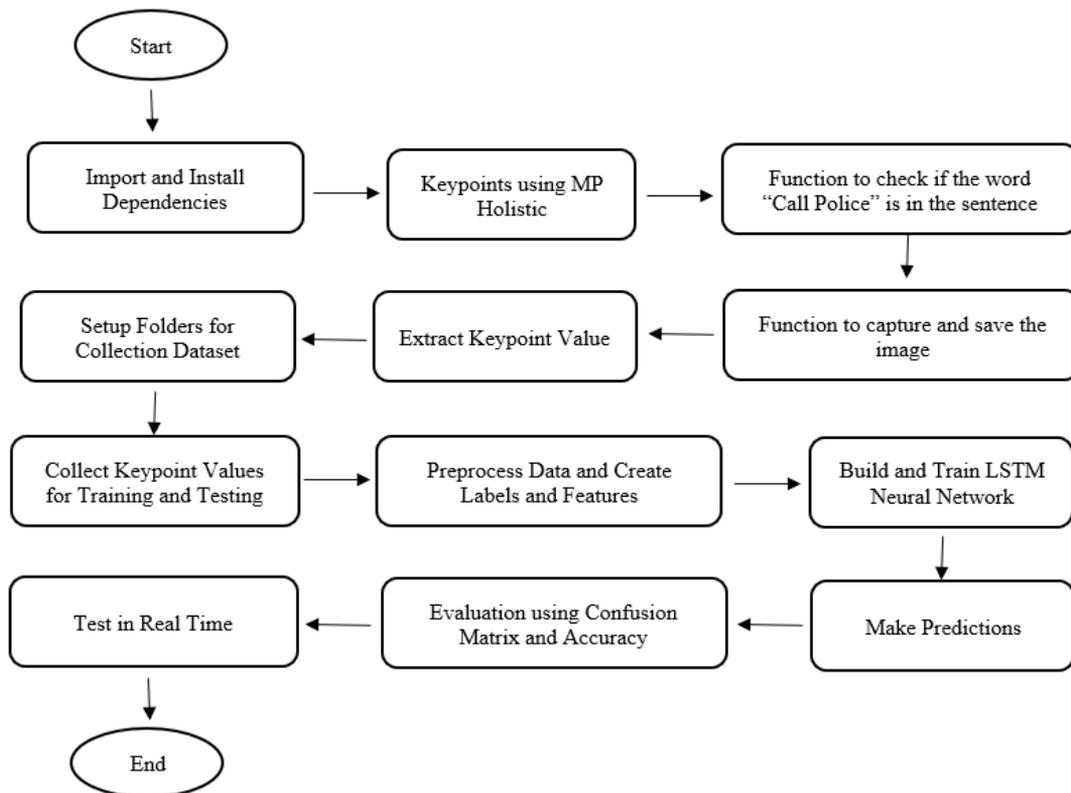


Figure 4. Flow chart depicting the methodology

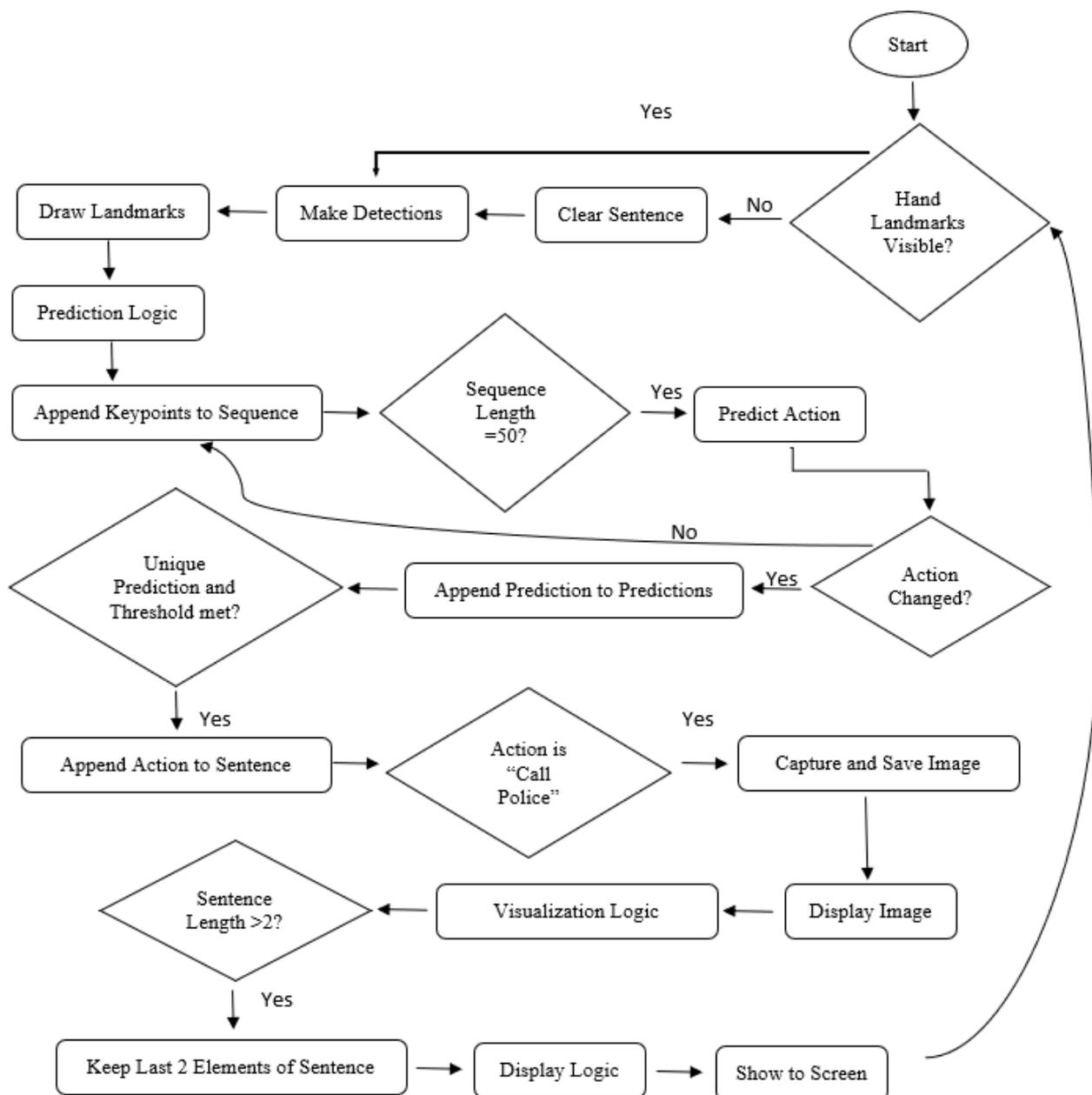


Figure 5. Flowchart depicting the system's implementation

- Real-time Detection: Once trained, the LSTM model can analyze incoming sequences of hand movements in real time.

- Response mechanisms: When violent hand signs are detected, the system may trigger pre-defined.

The actual implementation of the project code for building a violence sign recognition system based on deep learning for two levels involved using landmark points to detect the hand and using LSTM to recognize hand gestures. The main reason for using the LSTM algorithm is because of the dynamic relationship between the gesture movements, which reflects the relationship between the sequential frames. Since the algorithm works to save the previous state and give the decision after the new state arrives, it was the most suitable choice for detecting the assistive gestures, which are based on three important sequential movements. It will cover things like libraries and methodologies used in the development process. The code snippets, algorithms, and techniques used to achieve the project objectives will be detailed here.

This pseudocode outlines the main steps and functions of the code in a concise format:

1. **Initialize MediaPipe Solutions:**
 - Assign instances of MediaPipe classes to variables for holistic model and drawing utilities:
 - `mp_holistic = mp.solutions.holistic`
 - `mp_drawing = mp.solutions.drawing_utils`
2. **MediaPipe Detection Function:**
 - Define a function `mediapipe_detection`:
3. **Draw Styled Landmarks Function:**
 - Define a function `draw_styled_landmarks`:
 - Draw styled connections for face, pose, left hand, and right hand on the image using detection results.
4. **Help Detection and Image Capture Functions:**
 - Define a function `is_help`:
 - Check if the phrase "Call Police" is present in the given sentence.
 - Define a function `capture_and_save_image`:
5. **Extract Keypoints Function:**
 - Define a function `extract_keypoints`:
 - Extract keypoints from the detection results.
6. **Model Definition and Training:**

- Define the model architecture.
 - Compile the model with specified optimizer and loss functions.
 - Train the model on the training data.
7. **Action Prediction:**
 - Use the trained model to predict actions for the test data.
 - Print the predicted and true actions.
 8. **Save Model Weights:**
 - Save the trained model weights for future use.
 9. **Evaluate Model Performance:**
 - Evaluate the model's performance using metrics such as the confusion matrix and accuracy.
 10. **Real-time Detection and Action Handling:**
 - Predict the action.
 - Update predictions and form a sentence.
 - Capture and save the image if the predicted action is "Call Police".
 - Display the action sentence.

Detection variables are initialized, and a loop is implemented to continuously process frames from the camera feed. Within the loop, detections are made using the Mediapipe model, styled landmarks are drawn on the image, actions are predicted, and relevant actions trigger further actions such as capturing and saving images if "Call Police" is detected. The loop continues until a quit command is issued.

4. RESULT AND DISCUSSION

Propose a hand gesture recognition system using LSTM algorithms to facilitate communication in emergencies. Achieve high accuracy in real-time gesture detection to ensure prompt assistance during emergencies. Investigate the selection of good methods to detect hand gestures within various states, like illumination variation, rotation, and scaling, based on the shapes of hands by using landmark points. The implementation of a real-time hand gesture recognition system using a long-term memory (LSTM) algorithm has yielded promising results in meeting urgent communication needs during violent situations. The system demonstrated remarkable efficiency, achieving an average detection accuracy of 95% in recognizing distress signals. By integrating deep learning and computer vision techniques, the system successfully captured and interpreted hand gestures, enabling individuals to immediately seek help without relying on verbal communication. Photographs depicting the execution of the project are visible below. In the figure below (Figure 6), we notice how frames of hand movements were collected through the computer camera that express the words "help" and "call the police." 2,500 were collected for each gesture. Thus, we obtained a dataset with a total of 5000 images.

Now, an examination will be conducted on his findings related to manual gestures. It will be observed that he identified the gesture signifying the term "help" and also discovered the word "call police." After he discovers the word "call police," he will take a snapshot, save it in a folder, and then display it as shown in Figure 6.

For this study, we utilized specific gestures, namely "Help" and "Call the police," which were customized to align with the research requirements.

Likewise, it can be seen that the system is able to recognize

gestures and clearly detect body as the distance from the individual to the camera increases. Furthermore, the gesture detection accuracy is consistent, as shown in (Figure 7).



Figure 6. System implementation

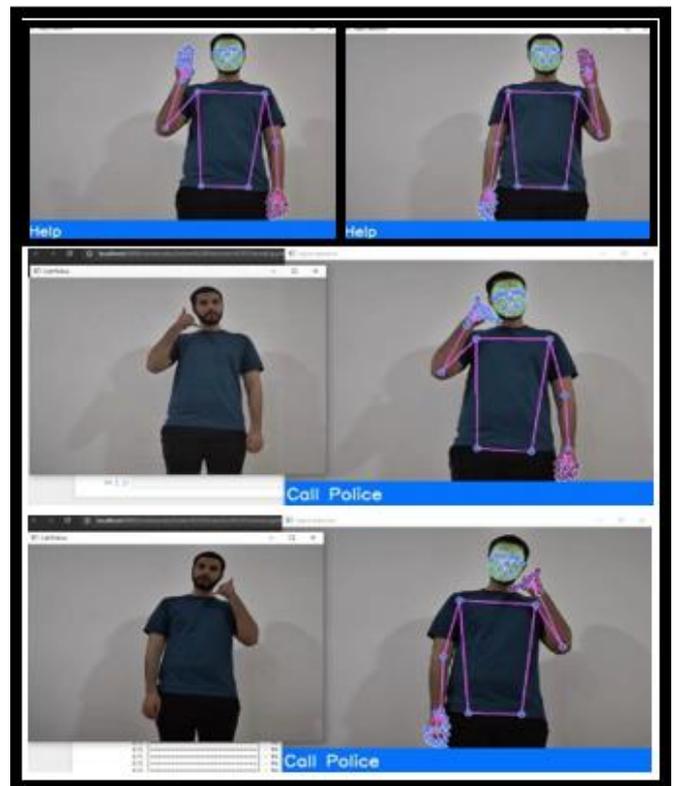


Figure 7. Executing the system from a greater distance

Accuracy is the most common measure for estimating the efficiency of the system. In this research, two types of dynamic gestures were identified: the first is the gesture of requesting help that spread during the Corona health crisis, and the second gesture is the gesture of requesting the police. Both gestures were entered into the system separately during the training process, and both gestures obtained an accuracy of 95. we notice how frames of hand movements were collected through

the computer camera that express the words “help” and “call the police.” 2,500 were collected for each gesture. Thus, we obtained a dataset with a total of 5000 images sequentially.

It is worth noting that the image on the left is the result of the system, taking a snapshot of the victim after identifying the help sign and call the police. The system also works on both the right and left hands, as shown in the figure above.

5. CONCLUSIONS

This research delves into dynamic hand gesture recognition and body tracking, leveraging deep learning methodologies, particularly the Long Short-Term Memory (LSTM) algorithm with two parameters and 2000 epochs used in research. The urgency and necessity for effective communication tools in critical situations, such as violence, accidents, and disasters, are underscored. Hand gestures emerge as a potent alternative, providing a lifeline for individuals ensnared in perilous circumstances. The convergence of artificial intelligence and real-time gesture recognition systems, epitomized by the LSTM algorithm, promises a beacon of hope, offering swift and discrete avenues for seeking assistance. We conclude that background and illumination have a significant impact on the recognition process, and the use of landmark points and skeleton points has greatly contributed to improving recognition and tracking in different environments. Using the LSTM algorithm requires a large and sufficient data set for training, and LSTM gave promising results in recognizing many of the gestures used in this research. Real-world scenarios demonstrate the effectiveness of the LSTM algorithm. With an accuracy rate soaring to 95%, the system showcases its prowess in swiftly identifying distress signs and orchestrating timely interventions. Moreover, the integration of body tracking mechanisms amplifies the system's efficacy, ensuring precision in location and movement tracking.

Subsequent investigations ought to concentrate on broadening the dataset to encompass a more diverse array of hand movements, body kinds, and environmental circumstances. This will enhance the gesture recognition system's accuracy and resilience in a range of real-world situations.

The proposed system does not pose any privacy concerns, especially since it can be applied through surveillance cameras in the streets and other places. The tracking process is only done if the victim herself requests help by the help or call police gestures.

REFERENCES

- [1] Chang, V., Eniola, R.O., Golightly, L., Xu, Q.A. (2023). An exploration into human-computer interaction: Hand gesture recognition management in a challenging environment. *SN Computer Science*, 4(5): 441. <https://doi.org/10.1007/s42979-023-01751-y>
- [2] Fadel, N., K. Abbood, I., Qasem Ghani, H. (2022). Best classification of continuous data based on hybrid decision tree. *International Journal of Intelligent Systems and Applications in Engineering*, 10(1s): 388-392.
- [3] Fadel, N., Kareem, E.I.A. (2022). Computer vision techniques for hand gesture recognition: Survey. In *International Conference on New Trends in Information and Communications Technology Applications*, Baghdad, Iraq, pp. 50-76. https://doi.org/10.1007/978-3-031-35442-7_4
- [4] Wu, W., Shi, M., Wu, T., Zhao, D., Zhang, S., Li, J. (2019). Real-time hand gesture recognition based on deep learning in complex environments. In *2019 Chinese Control and Decision Conference (CCDC)*, Nanchang, China, pp. 5950-5955. <https://doi.org/10.1109/CCDC.2019.8833328>
- [5] Thejowahyono, N.F., Setiawan, M.V., Handoyo, S.B., Rangkuti, A.H. (2022). Hand gesture recognition as signal for help using deep neural network. *International Journal of Emerging Technology and Advanced Engineering*, 12(2): 37-47. https://doi.org/10.46338/ijetae0222_05
- [6] Fadel, N., Kareem, E.I.A. (2022). Detecting hand gestures using machine learning techniques. *Ingenierie des Systemes d'Information*, 27(6): 957-965. <https://doi.org/10.18280/isi.270612>
- [7] Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.L., Grundmann, M. (2020). Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214*. <https://doi.org/10.48550/arXiv.2006.10214>
- [8] Liu, H., Zhang, H., Mertz, C. (2019). DeepDA: LSTM-based deep data association network for multi-targets tracking in clutter. In *2019 22th International Conference on Information Fusion (FUSION)*, Ottawa, ON, Canada, pp. 1-8. <https://doi.org/10.23919/FUSION43075.2019.9011217>
- [9] Pavlikov, A., Volkogonov, V., Lipatova, A. (2024). On the application of human pose estimation in a driver condition monitoring task. In *2024 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, Russian Federation, pp. 1-6. <https://doi.org/10.1109/IEEECONF60226.2024.10496788>
- [10] Sharma, S., Sudharsan, B., Narahariseti, S., Trehan, V., Jayavel, K. (2021). A fully integrated violence detection system using CNN and LSTM. *International Journal of Electrical & Computer Engineering*, 11(4): 3374-3390. <https://doi.org/10.11591/ijece.v11i4.pp3374-3380>
- [11] Toro-Ossaba, A., Jaramillo-Tigeros, J., Tejada, J.C., Peña, A., López-González, A., Castanho, R.A. (2022). LSTM recurrent neural network for hand gesture recognition using EMG signals. *Applied Sciences*, 12(19): 9700. <https://doi.org/10.3390/app12199700>
- [12] Omarov, B., Narynov, S., Zhumanov, Z., Gumar, A., Khassanova, M. (2022). State-of-the-art violence detection techniques in video surveillance security systems: A systematic review. *PeerJ Computer Science*, 8: e920. <https://doi.org/10.7717/peerj-cs.920>
- [13] Taskiran, M., Killioglu, M., Kahraman, N. (2018). A real-time system for recognition of American sign language by using deep learning. In *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, Athens, Greece, pp. 1-5. <https://doi.org/10.1109/TSP.2018.8441304>
- [14] Mujahid, A., Awan, M.J., Yasin, A., Mohammed, M.A., Damaševičius, R., Maskeliūnas, R., Abdulkareem, K.H. (2021). Real-time hand gesture recognition based on deep learning YOLOv3 model. *Applied Sciences*, 11(9): 4164. <https://doi.org/10.3390/app11094164>
- [15] Malik, N.U.R., Sheikh, U.U., Abu-Bakar, S.A.R., Channa, A. (2023). Multi-view human action recognition

- using skeleton based-FineKNN with extraneous frame scrapping technique. *Sensors*, 23(5): 2745. <https://doi.org/10.3390/s23052745>
- [16] Zhang, Z., Liu, S.L., Wang, Y.Y., Song, W., Zhang, Y.H. (2024). Online cross session electromyographic hand gesture recognition using deep learning and transfer learning. *Engineering Applications of Artificial Intelligence*, 127(Part A): 107251. <https://doi.org/10.1016/j.engappai.2023.107251>
- [17] Yu, R., Gu, G.Y. (2023). Deep transfer learning-based adaptive gesture recognition of a soft e-skin patch with reduced training data and time. *Sensors and Actuators A: Physical*, 363: 114693. <https://doi.org/10.1016/j.sna.2023.114693>