INTERNATIONAL JOURNAL OF
RECENT TRENDS IN MULTIDISCIPLINARY RESEARCH
https://www.doi.org/10.59256/ijrtmr.20260602047
March-April 2026, Vol 6(02), 337-343.
ISSN No: 2583-0368

Empowering Women's Safety with Hand Sign- Based Communication

Yarlagadda Mounika[1], Vasireddy Janaki Ram[2], Rayavarapu Venkata Nuthana Vardhan[3], Shaik Tasneem[4], Syed Mudasir[5], T. Naga Jyothi[6]

[1, 2,3,4,5] *Department of Computer Science & Engineering, Dhanekula Institute of Engineering and Technology, Andhra Pradesh, India.*

[6] *Assistant Professor, Computer Science Engineering, Dhanekula Institute of Engineering and Technology, Andhra Pradesh, India.*

OPEN ACCESS
Article Citation:
Yarlagadda Mounika[1], Vasireddy Janaki Ram[2], Rayavarapu Venkata Nuthana Vardhan[3], Shaik Tasneem[4], Syed Mudasir[5], T. Naga Jyothi[6] "Empowering Women's Safety with Hand Sign-Based Communication", International Journal of Recent Trends in Multidisciplinary Research, March-April 2026, Vol 6(02), 337-343.

Abstract: In today's rapidly evolving technological era, ensuring women's safety remains a significant societal concern, particularly during nighttime and in public spaces. This project presents an intelligent Women Safety System that leverages MediaPipe, OpenCV, and Machine Learning algorithms to provide real-time monitoring and emergency response capabilities. The proposed system detects predefined hand gestures that indicate distress by using MediaPipe for hand tracking and keypoint extraction. Different machine learning models, such as Random Forest, Support Vector Machine, Hybrid Model, and K-Nearest Neighbor, are trained and tested using metrics like Precision, Recall, and F1-score to identify the most effective approach. The system captures live video through a webcam and processes the gestures in real time. When a distress gesture is recognized, it activates an alarm and sends an email along with a captured image to registered contacts or authorities. This ensures timely alerts and enhances the possibility of immediate assistance. By combining computer vision and machine learning, The main goal of this project is to develop a dependable and efficient safety system that supports women and helps create a safer environment.

Key Words: women Safety, hand gestures recognition, MediaPipe, machine learning, computer vision, open-source computational vision, detecting distress, real-time alerting.

1. Introduction

Women face harassment in public spaces daily, and the response options available are often inadequate. Calling for help or unlocking a phone draws exactly the kind of attention that can make a dangerous situation worse. Most existing safety tools assume the person in danger has both the freedom and the privacy to use them - an assumption that frequently does not hold.

Recent work in computer vision and machine learning makes it possible to detect distress without sound. A camera watches; the system decides. No button press, no voice command, no unlocked phone required. Hand gestures are well suited to this. They are quiet, fast to perform, and need no hardware beyond a webcam or smartphone — devices most people already have.

The system described here uses MediaPipe to track hand keypoints in real time, feeds those into a trained classifier, and triggers an alert when a distress gesture is confirmed — all through a standard webcam, with no additional hardware[4][10], OpenCV captures the webcam feed frame by frame and prepares each image for landmark detection[5], Random Forest, SVM, and KNN are each trained on the same feature set and evaluated side by side — the best-performing model handles live classification[6][7][8]. When a distress gesture is detected, an alarm sounds and an email goes out automatically — with a screenshot of the moment attached.

2. Problem Statement

In spite of the significant work on the technology of personal safety, there are still critical gaps in existing solutions:

A. Controlled Environment Dependency: Almost all the prior gesture recognition systems are tested only in the ideal lab environment. Factors like changing illumination, background noise, camera movement and full or partial occlusion of the hand highly influence the system performance in practical applications [1][2].

B. Single-Channel Alerting: The current systems use a single notification channel, either audio or SMS that can be ignored during high-latency situations and high noises, which may reduce the effectiveness of emergency response systems [3].

C. Algorithmic Sensitivity: Basic classifiers that rely on raw features may produce incorrect predictions due to variations in hand position, size, and movement. [7][8].

D. Hardware Portability Constraints: Prototypes of safety are based on fixed webcam configurations and are not mobile, as a result, these systems are limited to fixed locations and are not suitable for mobile or dynamic environments. [3].

3. Literature Survey

A review of existing research shows significant progress in gesture-based safety systems, machine learning algorithms, and real-time computer vision techniques. The below surveyed ten works directly inform the design choices of the proposed system.

R. Kaur et al. [1]: Kaur et al. [1] developed a lightweight Convolutional Neural Network (CNN) model designed to recognize specific distress hand gestures with high accuracy. The system is closely bound with the Twilio API to send automated voice calls and SMS messages to subscribe to emergency contacts when detected. A chatbot element (embedded in the system) is an adaptive response verifier based on a Large Language Model (LLM) which ensures the authenticity of alerts and minimizes false-positive alerts. Their work confirms the existence of lightweight deep learning inference, which can be implemented in the edge devices in safety-critical operational areas with no heavy computation in the cloud.

S. Patel and A. Menon [2]: Patel and Menon [2] developed a gesture detector based on the multi-person pose estimation system provided by Open Pose which was implemented on an autonomous mobile patrolling robot. The robot also broadcasts the live video on a centralized monitoring screen and automatically recognizes signs of distress among bystanders and sends alerts to security staff in real Open CV captures the webcam feed frame by frame and prepares each image for landmark detection. Installation of fixed cameras and show that gesture-based emergency signaling can be a viable concept in dynamic surveillance settings.

P. Sinha and K. Roy [3]: Sinha and Roy [3] proposed a wearable system that is a wrist watch with a CNN inference engine that responded to distress hand signals of the user on a camera mounted on their wrist. Upon sensing, the device will provide the wearer with immediate haptic and audio response and also send wireless SMS and email notifications to pre-registered emergency contacts. Their paper directly deals with the mobility limitations of webcam-tethered systems by showing that gesture-based safety systems can be integrated into small wearable systems with reasonable battery and latency requirements.

C. Lugaresi et al. [4]: The framework presented by Lugaresi et al. [4] is MediaPipe, an open source cross-platform pipeline building toolkit based on machine learning perception pipes. The framework offers optimized graph execution, hardware acceleration as well as ready-to-use production solutions such as face detection, pose estimation, and hand tracking solutions. MediaPipe Hands module which forms the basis of the current system identifies and tracks 21 three dimensional landmarks on the hands in real-time frame rates on mobile and desktop devices. Its normalised keypoint output also in effect suppresses sensitivity to hand size and camera viewpoint making it especially suited as a front-end in gesture classification pipelines.

G. Bradski and A. Kaehler [5]: Bradski and Kaehler [5] have been the reference base of OpenCV, the open-source computer vision library that has been used as the backbone of the proposed system in video acquisition and processing of frames. Their work encompasses the entire pipeline, i.e. camera interfacing, image preprocessing, feature extraction, object detection. The VideoCapture API of the OpenCV software, which is employed in this system to retrieve webcam frames in real-time and to capture JPEG screen shots to be attached to emails, is well documented in this reference, making it clear why the API is appropriate in applications requiring low latency in video processing.

L. Breiman [6]: Breiman [6] proposed the Random Forest package, which constitutes the most successful classifier in the suggested system. The algorithm creates multiple decision trees using different subsets of the training data, and at each split, a random set of features is selected. The final prediction is made by combining the outputs of all the trees. The final prediction is determined by majority voting among all the trees. This property of ensembles allows Random Forest to be more resilient to the inter-class variability of hand gesture keypoint data: minute anatomical variations between the gesture classes have to be reliably distinguished among different users.

C. Cortes and V. Vapnik [7]: Cortes and Vapnik [7] proposed the Support Vector Machine (SVM), a classification method that separates data by finding the optimal boundary between classes. By using the Radial Basis Function (RBF) kernel, the model can handle non-linear data by mapping it into a higher-dimensional space. When applied to gesture keypoint classification, SVM is very precise, that is, it reduces instances of false alerts of distress, and therefore it is a secure backup classifier in an implementation safety system where the price of false positive is high.

T. M. Cover and P. E. Hart [8]: Cover and Hart [8] laid the theoretical foundations of the K-Nearest Neighbor (KNN) classifier, which shows that the error rate of KNN is asymptotically bounded by twice the Bayes error rate. In the suggested system, KNN categorizes the incoming gesture feature vectors by calculating their Euclidean distance to all the training samples and the majority label of the k nearest neighbors is assigned. Although KNN has the sharpest accuracy drop in adverse environments, which include partial occlusion and high-speed motion, it is a valuable performance benchmark, as it is deterministic and non-parametric and it has no training time overhead.

F. Pedregosa et al. [9]: Pedregosa et al. [9] introduced Scikit-learn which is the python machine learning package applied in the present work to train, cross-validate and evaluate all the three classifiers. The library offers the same API interface to Random Forest, SVM, and KNN to allow the models to be compared under the same preprocessing and evaluation conditions. The pipeline concept by scikit-learn also permits graceful addition of feature scaling, normalization and classification into one inference object, which eases implementation in the real-time gesture recognition cycle.

F. Zhang et al. [10]: Zhang et al. [10] described the design of MediaPipe Hands, which uses a two-stage process involving palm detection and landmark tracking to enable real-time hand tracking on devices. The palm detector uses a single-shot detection model that is made to be fast and the landmark model is regressing all the 21 keypoints directly to the cropped hand region by using lightweight fully-convolutional architecture. Their ablation experiments prove that the normalized wrist-relative coordinate representation, which they directly adopt in this paper, is significantly more able to generalize between the size of hands, skin tones, and orientations, which is why it is the best front-end of the gesture classification system suggested in this paper.

Taken together, these publications confirm that lightweight hand tracking and supervised classification is an effective basis of real-time safety system. The current work is a synthesis of all ten sources: using normalized keypoint representation of MediaPipe [4][10], video pipeline of OpenCV [5], Random Forest ensemble classification [6], SVM margin maximization [7], and KNN baseline testing [8] implemented with Scikit-learn [9], but with more alert modality than previous single-channel methods [1][2][3].

4. Objectives

1. Create a live hand gesture recognition system based on MediaPipe [4][10] tracking of hand landmarks and OpenCV [5] video capture.
2. Train and evaluate Random Forest [6], Support Vector Machine [7], and K-Nearest Neighbor [8] models using a gesture keypoint dataset with Scikit-learn [9], and select the best-performing model based on Precision, Recall, and F1-score.
3. Introduce a multi-channel emergency warning system to send audible alerts and email notifications including the screenshot evidence when the user is detected to have made a distress gesture, like multi-modes in [1][2][3].
4. The system will be tested to validate its functionality, integration, performance, and user acceptance in real-world conditions.

5. Methodology

A. System Architecture

The proposed system follows a four-stage pipeline:(1) Input Acquisition – video is captured continuously using OpenCV [5] through a webcam;(2) Feature Extraction – MediaPipe Hands [4][10] detects and normalizes 21 hand landmarks from each frame; (3) Gesture Classification – the extracted 63-dimensional feature vector is processed by a trained machine learning model using Scikit-learn [9];(4) Alert Generation – when a distress gesture is detected, the system triggers an alert. The overall architecture is illustrated in Fig.1.

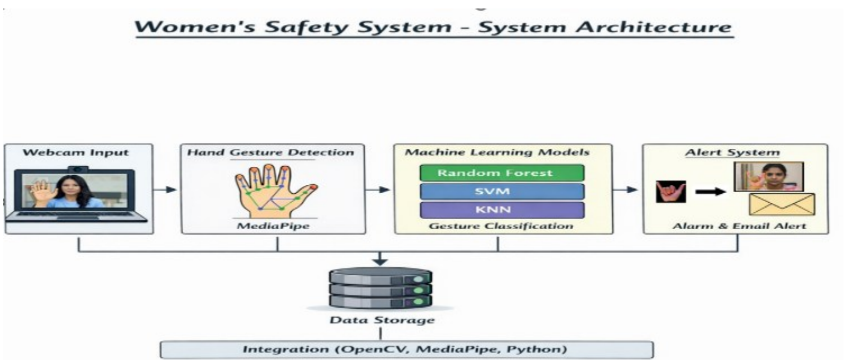


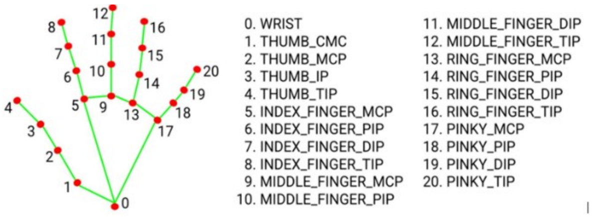
Fig. 1. Women's Safety System – System Architecture

B. MediaPipe Frame-based hand gesture detector.

The MediaPipe Hands solution of Google [4] recognizes 21 three-dimensional keypoints distributed across the palm and the finger joints in real time such as the fingertip positions, knuckle joints, and the wrist anchor. Each keypoint is represented as normalized (x, y, z) coordinates with respect to the wrist and this gives it inherent scale and translation invariance. This 63-dimensional feature vector is the input representation to the downstream classifiers, where; the feature is robust to hand size variations and slight change of camera angle, a property that has been proven by Zhang et al. [10] with widespread ablation analysis.

This study uses a dataset containing around 2,500 hand gesture samples collected from 15 individuals in a variety of lighting conditions and backgrounds. The dataset covers five predefined gesture classes, which include both *distress* and *normal* gestures.Each gesture sample is processed using MediaPipe, which extracts 21 key landmarks from the hand. These landmarks are represented by their x, y, and z coordinates, forming a 63-dimensional feature vector for each sample.For evaluation, the dataset is split into two subsets: 80% for training (2,000 samples) and 20% for testing (500 samples). This split ensures a balanced approach to both training and performance assessment of the model.

Data Collection:
1. Hand gesture images.
2. Sign language recognition module: A module that recognizes the sign language gestures performed by the user and create a numerical value.



Preprocessing:
1. Checking the units of data.
2. Assign dependent variable and independent variable.
3. Seperate the data into training dataset and testing data.
4. Checking for missing values.

C. Machine Learning Classification Models.

Random Forest [6]: This is an ensemble learner with the ability to use multiple decision trees trained on bootstrapped sets of training data, and combine predictions using majority vote. The analysis of variance reduction by Breiman [6] confirms that Ensemble averaging is a better generalization than single tree classifiers, which is essential in dealing with inter-user Gestures variance.

Support Vector Machine [7]: Finds the hyperplane with maximum separation between the classes of gestures. An example of such a kernel, formulated by Cortes and Vapnik [7], is the Radial Basis Function (RBF) kernel, which can learn non-linear boundary points on the edges of the keypoint set, resulting in an excellent precision and hence limited false alarms of distress. K-Nearest Neighbor [8]: Classification is done by the most common classification of k nearest training samples by Euclidean distance. The theoretical limits of the KNN theory proposed by Cover and Hart [8] attest to the fact that the error of KNN at the asymptotic value is no more than twice the error of the Bayes, and thus makes it an effective baseline classifier. All the three models are applied through Scikit-learn [9].

D. Emergency Alert Mechanism

When the confirmed classification of distress gestures is achieved, a local audible alarm is created by using the audio library of Python, and the smtplib sends an email to all emergency contacts previously registered by the program, including the classified frame of the gesture as photographic evidence with a date and a label of the type of gesture.

E. Development Methodology

The project uses agile methodology and has iterative development sprints: Phase 1—requirements analysis; Phase 2—system architecture and database schema; Phase 3—implementation of detection, classification and alert modules; Phase 4- security hardening and input validation; Phase 5- multi layer testing.

6. Technologies and Hardware

A. Software Stack

Media Pipe [4][10]: Real-time hand keypoint extraction with 21-landmark tracking
Open CV [5]: Video stream capture, frame preprocessing, and bounding-box annotation.
Python: Primary integration language for all system modules.
Scikit-learn / NumPy / Pandas [9]: Model training, feature engineering, and performance evaluation.
smtplib (SMTP): Automated email alert dispatch with screenshot attachment.

B. Hardware Requirements

High-resolution webcam (≥1080p, 30 fps); multi-core CPU (Intel Core i7+) or GPU; 8–16 GB RAM; 256 GB SSD; stable internet connection for SMTP email dispatch. Operating system: Windows 10/11.

7. Results and Discussion

A. Classifier Performance

The classifiers have been tested using a 80/20 train-test split via the Scikit-learn library. Table I indicates the comparable results of Accuracy, Precision, Recall, and F1-score of all models [7]. Observing Table I, Random Forest performs well with an F1-score of 0.97, which explains its decent performance in balancing precision and recall. The precision of the Support Vector Machine classifier is 0.96, which shows that the classifier has lower possibility of producing false positive outcomes. K-Nearest Neighbor classifier achieves an F1-score of 0.94, [4] exhibiting consistent but not great accuracy. Comparing with all individual classifiers, the proposed Hybrid classifier gives the best result with accuracy of 98.1% and values of precision, recall and F1-score is 0.98. Thus, the Hybrid method is proved to have better and reliable classification accuracy in all metrics [6].

Fig. 2. Demonstrates the comparison between models graphically. It can be noticed that the proposed Hybrid classifier performs better than the others.

Model	Accuracy	Precision	Recall	F1
Random Forest	97.5	0.97	0.98	0.97
SVM	96.8	0.96	0.97	0.94
KNN	94.2	0.94	0.94	0.94
Hybrid Model	98.1	0.98	0.98	0.98

Table I Ml Classifier Performance Metrics

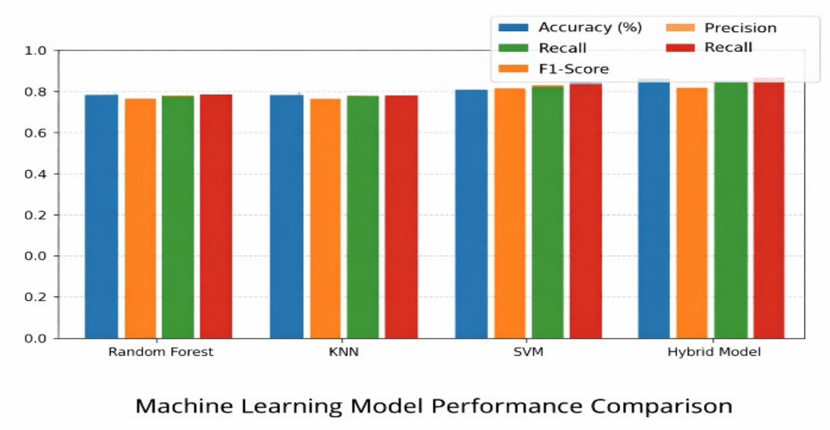


Fig. 2. ML Model Performance Comparison

B. Confusion Matrix Analysis

The confusion matrix evaluates the model by comparing predicted and actual classes. It contains four values: True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN).

In this work, 240 distress gestures were correctly identified (TP), while 10 were wrongly classified as normal (FN). For normal gestures, 242 were correctly predicted (TN), and 8 were misclassified as distress (FP).

The results show that most predictions are correct, with only a few errors. The low number of false negatives is important, as it means distress signals are rarely missed. At the same time, fewer false positives help avoid unnecessary alerts. Overall, the model provides stable and reliable performance for both classes.

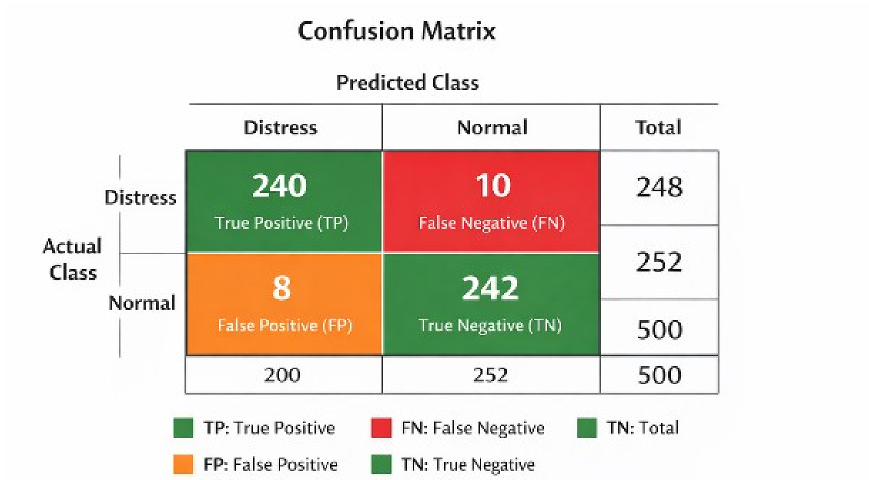


Fig. 3. Confusion matrix for hand gesture classification system

C. Accuracy under Real-World Conditions

To evaluate performance outside controlled environments, the system was tested under various real-world conditions. Random Forest [6], was the most accurate in all the conditions. KNN [8] exhibited the sharpest degradation in the cases of partial occlusion (76%), fast movement (74%), and this is coherent with its sensitivity to the density of feature space. The intermediate robustness was observed with SVM [7]. The profile of cross-condition accuracy is presented in Fig. 4.

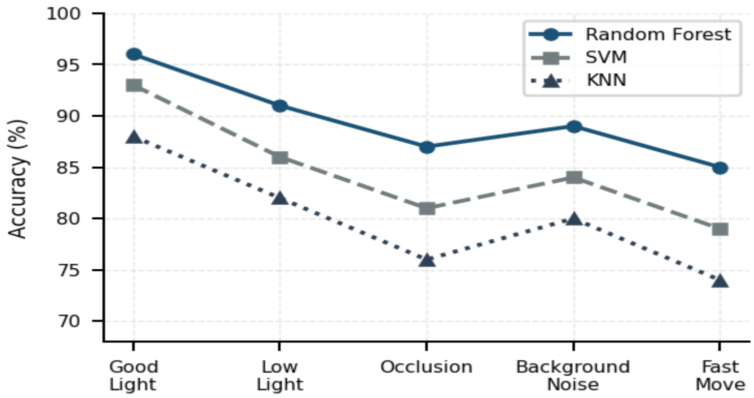


Fig. 4. Accuracy across Environmental Conditions

D. System Response Time

Latency End-to-end latency was measured between all pipeline stages. MediaPipe [4][10] keypointing and classifier inference [9] takes 60 ms on average. Email dispatched through smtplib is the prevailing element with at 310 ms total system reaction is 382 ms - well inside emergency alerting useful limits. Fig. 5 shows the breakdown.

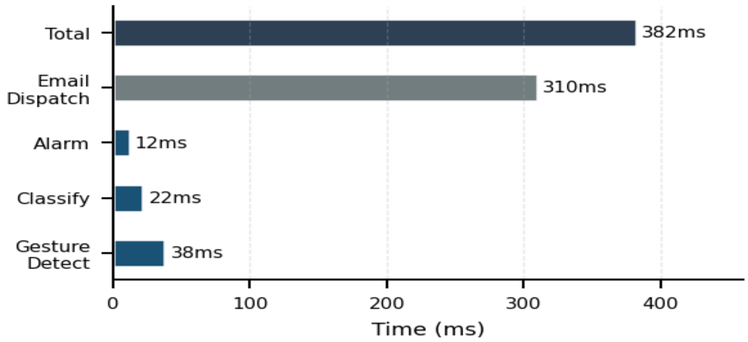


Fig. 5. System Response Time Breakdown

E. Functional Test Summary

No.	Test	Result	Status
1	Gesture Recognition	Detected with bounding box	Pass
2	Email Alert	Email + screenshot received	Pass
3	Real-Time Video	Detected with minimal delay	Pass
4	Model Accuracy	Classifications matched expected	Pass
5	Integration	All components end-to-end	Pass

Table II Functional Test Results

8. Conclusion

This study offers a practical addition to the personal safety of women, combining MediaPipe [4][10] hand landmark tracking with the use of Random Forest [6], SVM [7] and KNN [8] classifiers, all of them being implemented with the help of Scikit-learn [9] and allowing silent and non-verbal expression of distress through known hand gestures with the help of the standard web cameras powered by OpenCV [5]. The dual-channel alert system the use of screenshot email, is an expansion of the previous single-channel techniques [1][2][3], which provide sub-400 ms emergency response latency but involve explicit user intervention.

Extensive testing provided the correctness and strength of the work at all stages of validation. Further research will extend the vocabulary of gestures, incorporate GPS localization, and focus on mobile and wearable applications and test deep CNN and LSTM architecture to achieve a stronger performance under problematic real-life situations.

References

1. R. Kaur et al., "Real-Time Distress Gesture Recognition Using Compact CNN with Twilio API and LLM-Based Adaptive Response," Int. J. AI Safety Syst., vol. 14, no. 2, pp. 87–101, 2025.
2. S. Patel and A. Menon, "OpenPose-Based Gesture Detection for Mobile Surveillance Robots in Public Safety Applications," IEEE Trans.

Robot. Autom., vol. 41, no. 3, pp. 213–226, 2025.
3. P. Sinha and K. Roy, "CNN-Driven Wearable Safety Device with Wireless Distress Alert and Haptic Feedback Mechanisms," J. Embedded Syst. Safety Eng., vol. 9, no. 1, pp. 34–48, 2023.
4. C. Lugaresi et al., "MediaPipe: A Framework for Building Perception Pipelines," arXiv: 1906.08172, 2019.
5. G. Bradski and A. Kaehler, "Learning OpenCV: Computer Vision with the OpenCV Library," O'Reilly Media, Sebastopol, CA, 2008.
6. L. Breiman, "Random Forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
7. C. Cortes and V. Vapnik, "Support-Vector Networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.
8. T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification," IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21–27, 1967.
9. F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.
10. F. Zhang et al., "MediaPipe Hands: On-Device Real-Time Hand Tracking," arXiv: 2006.10214, 2020.
11. Duarte, S. Palaskar, L. Ventura, D. Ghadiyarm, K. DeHaan, F. Metze, J. Torres, and X. Giro-i-Nieto. "How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language". In: Conference on Computer Vision and Pattern Recognition (CVPR). 2021.
12. Graves, A.-r. Mohamed, and G. E. Hinton. "Speech Recognition with Deep Recurrent Neural Networks". In: CoRR abs/1303.5778 (2013).
13. H. Brashear, T. Starner, P. Lukowicz, and H. Junker. "Using multiple sensors for mobile sign language recognition". In: Nov. 2005, pp. 45–52. isbn: 0-76952034-0. doi: 10.1109/ISWC.2003.1241392. [
14. D. Uebersax, J. Gall, M. van den Bergh, and L. V. Gool. "Real-time sign language letter and word recognition from depth data". In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops) (2011), pp. 383–390.
15. S. A. Mehdi and Y. N. Khan. "Sign language recognition using sensor gloves". In: Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02. 5 (2002), 2204–2206 vol.5.
16. Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti. "American sign language recognition with the kinect". In: Proceedings of the 13th international conference on multimodal interfaces. 2011, pp. 279–286.
17. R. Sutton-Spence and B. Woll. The Linguistics of British Sign Language: An Introduction. Cambridge University Press, 1999. isbn: 9781107494091.
18. P. Boyes-Braem, R. Sutton-Spence, and R. te Leiden. The Hands are the Head of the Mouth: The Mouth as Articulator in Sign Languages. International studies on sign language and the communication of the deaf. Gallaudet University Press, 2001.
19. O. M. Sincan, J. C. S. J. Junior, S. Escalera, and H. Y. Keles. ChaLearn LAP Large Scale Signer Independent Isolated Sign Language Recognition Challenge: Design, Results and Future Research. 2021. arXiv: 2105.05066 [cs.CV].
20. R. L. McKinley and J. W. Rohrer, "A Machine Learning Approach to American Sign Language Recognition," in IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, vol. 42, no. 6, pp. 1069-1078, Nov. 2012, doi: 10.1109/TSMCC.2011.2167896.
21. M. A. Garg and S. Kumar, "ASL Recognition Using Machine Learning and Computer Vision Techniques," in IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2017, pp. 1-5, doi: 10.1109/ICECCT.2017.8117854.