Journal of Technology and Informatics (JoTI)
Vol. 8, No. 1, Bulan April Tahun 2026
P-ISSN 2721-4842
E-ISSN 2686-6102

A Real-Time Human-Drone Interaction System for Cornfield Perimeter Monitoring Using Hand Gesture Control

Fadzillah Akbar Subkhi[1*], Muhammad Fuad[2], Sri Wahyuni[3], Achmad Imam Sudianto[4], Vivi Tri Widyaningrum[5], Ach. Dafid[6]

[1*,2,3,4,5,6] Department of Mechatronics Engineering, Universitas Trunojoyo Madura, East Java, Indonesia
220491100042@student.trunojoyo.ac.id[1*], fuad@trunojoyo.ac.id[2], s.wahyuni@trunojoyo.ac.id[3], Aimam.sudianto@trunojoyo.ac.id[4], vivi@trunojoyo.ac.id[5], Ach.dafid@trunojoyo.ac.id[6]

Article Information

Article History:
Received : 12 December 2025
Revised : 30 January 2026
Accepted : 9 February 2026
Published : 27 April 2026

***Correspondence:**
220491100042@student.trunojoyo.ac.id

Keywords:
Drone Control, Gesture Recognition, MediaPipe Hands, Perimeter Monitoring, Smart Farming, Support Vector Machine (SVM)

10.37802/joti.v8i1.1314

Journal of Technology and Informatics (JoTI)
P-ISSN 2721-4842
E-ISSN 2686-6102
https://e-journals.dinamika.ac.id/index.php/joti

Abstract:

Perimeter monitoring in agricultural fields is essential for maintaining security and ensuring continuous observation of field conditions. This study develops a real-time human–drone interaction system using hand-gesture recognition based on MediaPipe Hands and a Support Vector Machine (SVM) classifier. A custom dataset of 24,000 images across 12 gesture classes was collected and converted into 42 hand landmarks (x, y, z), normalized relative to the wrist point. The SVM model with an RBF kernel was trained using an 80:20 split and achieved a testing accuracy of 99.18%. The system operates at 109 FPS with an average latency of 9.16 ms, enabling rapid and reliable drone responses to gesture commands. Field testing in a cornfield with FPV camera visualization demonstrated that the system consistently recognized gestures in varying outdoor lighting, allowing drones to execute precise perimeter checks and maneuvers. These results highlight the significant potential of integrating gesture recognition with drone control, providing a practical, real-world solution that advances smart farming, increases agricultural efficiency, and supports technological progress toward Sustainable Development Goals. The proposed system thus offers a lightweight, responsive, and impactful tool for modern agricultural perimeter monitoring.

INTRODUCTION

Perimeter monitoring in agricultural fields is essential for maintaining security, observing activities around crop areas, and ensuring that no disturbances occur that could reduce productivity [1][2]. In large cornfields, monitoring is generally still carried out manually by field operators, a method that requires considerable time and labor while offering limited visibility. Advances in drone technology have opened new opportunities for more efficient area surveillance [3] [4], as drones are capable of covering wider areas in relatively short periods. However, the

effectiveness of drone operations still depends heavily on the interaction method between the operator and the system. One increasingly adopted approach is the use of hand gestures as a more intuitive form of communication between humans and drones.

Various studies have explored hand gesture recognition for drone control using different approaches [5], including deep learning methods (techniques that allow systems to learn from data using neural networks) [6], motion sensors (devices that detect movement), and keypoint extraction–based systems (which identify and track specific points on the hand) [7]. Convolutional Neural Network (CNN) based approaches are popular due to their strong automatic feature extraction capabilities [8]. One study, for example, employed a CNN combined with HSV (Hue, Saturation, Value—a color-based image segmentation method) segmentation to recognize upper-body gestures for controlling a mobile robot [9]. Despite achieving high accuracy, CNN-based methods have notable limitations, particularly in outdoor environments [10]. These models typically require stable backgrounds, sufficient lighting, and higher computational resources. In cornfields, where lighting conditions are dynamic and visual elements are highly complex, such approaches become less optimal. Other research has demonstrated that drone control through gestures can be performed in simulated environments using Leap Motion (an optical hand-tracking sensor) and Unity (a game development platform) [11]. The results showed that while some static gestures were recognized well, dynamic gestures often suffered reduced accuracy due to sensor field-of-view limitations and hand occlusion.

Moreover, the system operated only within a simulation environment, without involving real drones or field conditions that demand rapid response and stable communication. Another study developed a dynamic gesture recognition system for UAV (Unmanned Aerial Vehicle) control using deep learning with a Leap Motion Controller [12]. Although the method achieved high accuracy under controlled conditions, it faced significant limitations. Leap Motion works optimally only indoors, requires a fixed operational range, and struggles with lighting variation and complex outdoor backgrounds [13]. The reliance on a specialized sensor also reduces system flexibility, making it unsuitable for open agricultural environments. Most previous studies have focused primarily on gesture recognition itself, rather than on the full integration of gesture systems, drone platforms, and field monitoring processes. In addition, only a limited number of works have examined how gesture-based systems respond to varying lighting conditions in outdoor agricultural settings. Research that specifically highlights human–drone interaction in real operational contexts, such as perimeter monitoring in cornfields, remains scarce.

The main objective of this study is to develop a human–drone interaction system [14], [15] based on hand gesture recognition using MediaPipe Hands (a framework for real-time hand tracking and keypoint detection) [14], [15], [16] and a Support Vector Machine (SVM) classifier (an algorithm for classifying data into different categories) for real-time perimeter monitoring in cornfields [17]. The system enables operators to control the drone directly through twelve predefined gestures without requiring a controlled background or ideal lighting conditions. Gesture detection is processed on a computer and transmitted to the drone to execute maneuvers such as arm, backward, disarm, down, forward, hover, landing, left, left_rotation, right, right_rotation, and up. This system is expected to support smart farming practices by improving labor efficiency, enhancing field security, and utilizing drones as real-time monitoring devices with a more intuitive form of interaction.

The subsequent sections of this manuscript are structured to provide a comprehensive overview of the proposed system: Section 2 details the research methodology, including dataset preparation, feature extraction via MediaPipe Hands, SVM-based classification, and the integration of the system with the drone platform. Section 3 presents a rigorous evaluation of the

experimental results and discussion, encompassing gesture recognition performance, real-time field testing, and a comparative analysis with existing literature. Finally, Section 4 concludes the study by synthesizing the core findings and outlining strategic directions for future research and system optimization.

METHOD

This section provides a comprehensive explanation of the research stages, beginning with the development of the gesture dataset, feature extraction using MediaPipe Hands, the design of the Support Vector Machine (SVM) classification model, and the integration of the gesture detection system with the drone for perimeter monitoring in cornfields.

Research Flow

This study employs an experimental approach in which the gesture recognition system is built and tested directly on a real drone operating in an actual cornfield environment. The overall workflow consists of four main phases: dataset collection, hand landmark extraction using MediaPipe, SVM model training, and real-time gesture detection integrated with drone control. The complete workflow of the proposed system is illustrated in Figure 1.

Gesture Dataset Collection

The gesture dataset was developed independently. Video recordings were converted into JPG images for each frame. The dataset contains 24,000 images that represent 12 gesture classes, with 2,000 samples for each. The gestures include: arm, backward, disarm, down, forward, hover, landing, left, left_rotation, right, right_rotation, and up. All gestures were recorded using two hands, and 21 landmarks were extracted per hand. The twelve gestures used for drone control are shown in Figure 2. The dataset was captured with a webcam at 30 FPS and stored in separate folders for each gesture class, as shown in Figure 3. This folder structure allows systematic labeling and simplifies supervised learning. Data collection was automated with a Python script to ensure cleanliness and minimize noise. This script verified the detection of all hand landmarks, as illustrated in Figure 4.

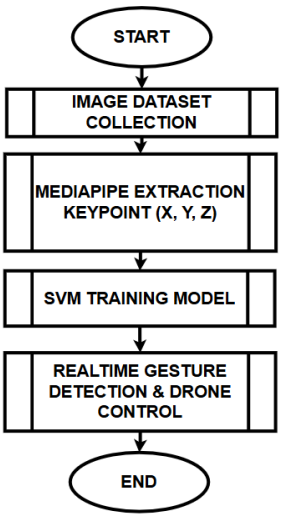

Figure 1. Research Workflow Flowchart

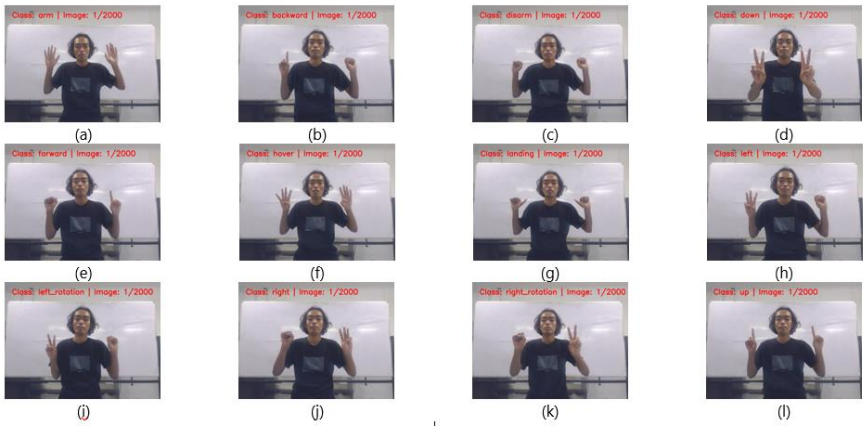

Figure 2. The Twelve Gesture Classes Representing (a) *arm*, (b) *backward*, (c) *disarm*, (d) *down*, (e) *forward*, (f) *hover*, (g) *landing*, (h) *left*, (i) *left_rotation*, (j) *right*, (k) *right_rotation*, dan (i) *up*

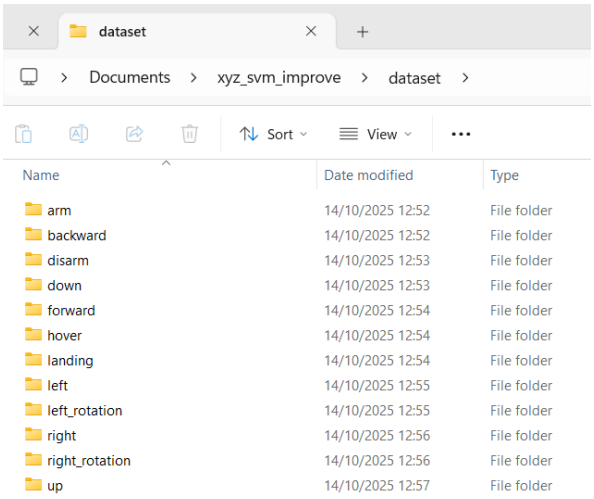

Figure 3. Dataset Folder Structure

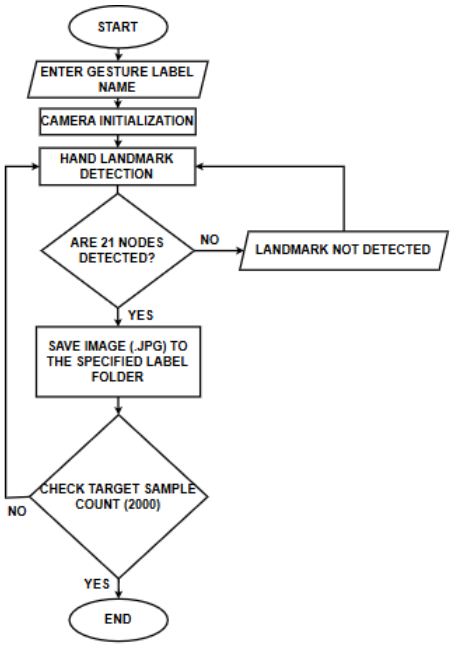

Figure 4. Dataset Acquisition Flowchart

MediaPipe Keypoint Extraction

Each landmark was normalized relative to the wrist point to ensure that variations in hand position and distance from the camera did not affect the extracted feature patterns. Feature extraction was performed using Google's MediaPipe Hands Detection framework [18], which detects hand landmarks from input images and converts them into three-dimensional coordinate data (x, y, and z). The spatial structure and indexing of the 21 hand landmarks detected by MediaPipe Hands are illustrated in Figure 5. This representation provides a clear reference for understanding the positional relationships among landmarks used as features in the classification process. Following landmark detection, the extracted coordinates were organized and stored in comma-separated values (CSV) files for further processing and model training. The overall workflow of the feature extraction process, from image input to normalized landmark output, is summarized in the flowchart shown in Figure 6.

Support Vector Machine Classification Model

The gesture classification model was developed using a Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel. The SVM algorithm was selected due to its high effectiveness in handling non-linear classification tasks and its robustness against overfitting in high-dimensional feature spaces [20]. Moreover, SVM offers lower computational cost and shorter training time compared to more complex deep learning models, making it suitable for real-time applications detection scenarios. The configuration parameters used during the training process are presented in Table 1.

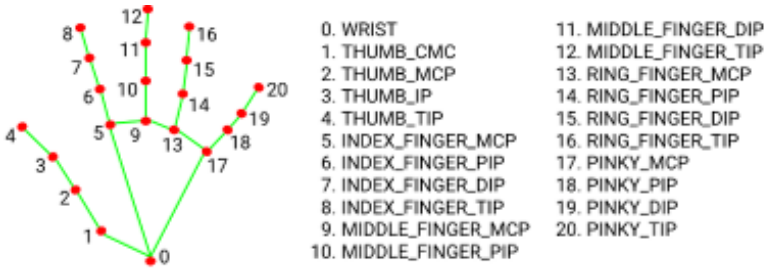

Figure 5. MediaPipe Hand Landmarks Coordinates Explanation [19]

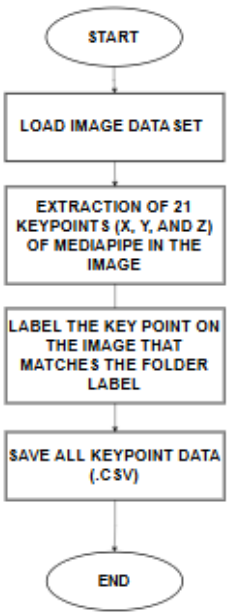

Figure 6. Feature Extraction Flowchart

Table 1. Parameter Configuration

Parameter	Description	Value
Kernel	Type of kernel function	RBF
C	penalty parameter	0.1
γ	Kernel coefficient	Scale
Input Features	3D coordinates (*x, y, z*) of 21 landmarks (both hands)	126 features
Scalling	Normalization method	Z-score normalization
Cross Validaton	Validation method for hyperparameter tuning	5-fold
Parameter Optimization	Search strategy	Grid search
Data Split	Train : test	80:20

For the RBF kernel, the coefficient γ was set to "scale" in the implementation, which automatically adjusts the value based on the number of features and the variance of the input data. The formulation used is shown (1):

$$\gamma = \frac{1}{n_{features}\, x\, Var(x)} \quad (1)$$

Description:

γ : *gamma*
$n_{features}$: number of input features
$Var(x)$: feature variance in the training set

Gesture Integration with the Drone

After the gesture classification process, the recognized gesture commands were transmitted from the laptop to the drone control system. The overall architecture of the gesture-to-drone integration, including data flow between the gesture recognition module and the drone control unit, is illustrated in Figure 7.

RESULTS AND DISCUSSION

This section presents the experimental results of the proposed system, focusing on gesture recognition performance and real-time field implementation. The SVM-based classifier achieved an accuracy of 99.18%, while the real-time system operated at an average speed of 109 FPS with a mean latency of 9.16 ms. These results indicate that the system is capable of recognizing hand gestures accurately and responding quickly during drone operation. The analysis is conducted at two levels: algorithmic performance and operational system performance under outdoor field conditions.

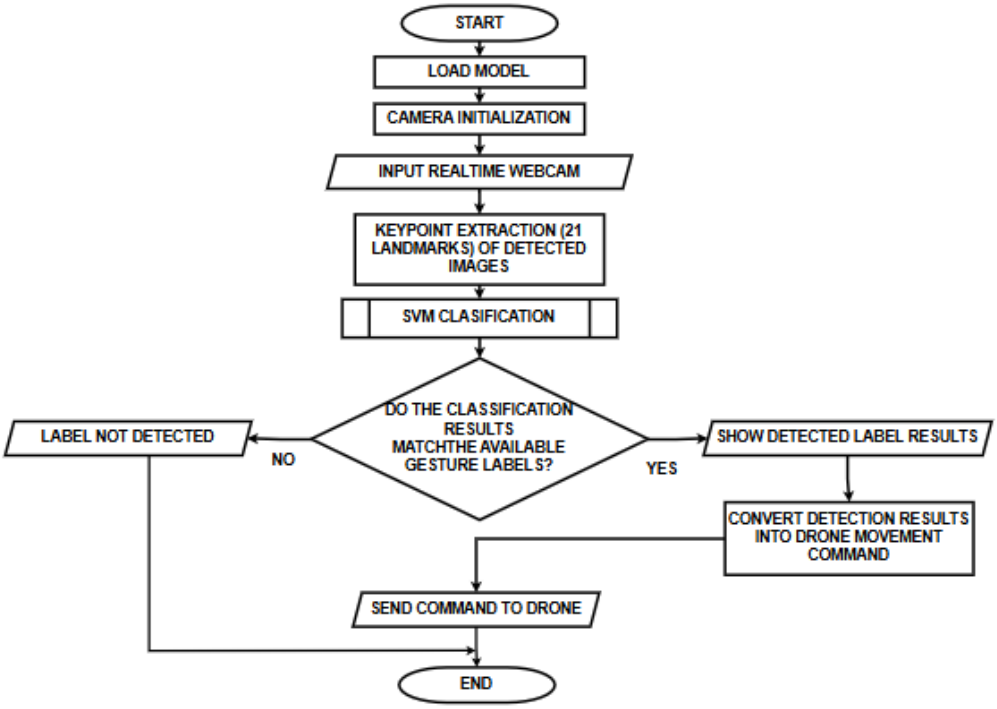

Figure 7. Flowchart Gesture Integration to Drone Control

Model Training Result

The SVM model with an RBF kernel was trained using a dataset of 24,000 normalized gesture images, where normalization was performed based on the relative distance to the wrist landmark. Of the total dataset, 80% was used for training and 20% for testing. Using the predefined parameters, the RBF-based SVM model achieved an accuracy of 99.18%. The model's performance in real-time detection was evaluated using precision, recall, and F1-score for each gesture label. These metrics are defined by the following equations (2), (3), and (4):

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall} \quad (4)$$

Description:

TP : True Positive

FP : False Positive

FN : False Negative

The resulting precision, recall, and F1-score for all gesture classes are summarized in Table 2.

Tabel 2. Clasification Report Realtime Detection

Gesture	Precision	Recal	F1-Score
arm	1.00	0.98	0.99
backward	0.99	0.94	0.99
disarm	0.98	1.00	0.99
down	0.94	0.99	0.97

Gesture	Precision	Recal	F1-Score
forward	0.99	0.99	0.99
hover	0.99	0.95	0.97
landing	0.99	0.98	0.99
left	0.99	1.00	0.99
left_rotation	1.00	0.99	0.99
right	0.99	0.99	0.99
right_rotation	1.00	0.99	0.99
up	0.99	1.00	0.99
Accuracy		0.9918	

Table 3. Real-Time Performance Gesture

Parameter	Result
Total sample	1420 frame
Average FPS	109.20 FPS
Average Detection (ms)	9.16 ms
Max Detection (ms)	83.83 ms
Min Detection (ms)	6.74 ms

Gesture Recognition Performance and Field Testing

Real-time inference was tested using a webcam under outdoor lighting conditions. The system achieved an average processing speed of 109 FPS with an average detection latency of 9.16 ms, as presented in Table 3. Examples of real-time detection using a webcam and the drone's perimeter monitoring view in the cornfield are summarized in Table 3. The experimental results obtained in the 8 × 10 meters cornfield under outdoor conditions are presented in Figures 8, 9, and 10. These figures illustrate the real-world implementation of the proposed system during field testing. Figure 8 presents the captured perimeter view of the cornfield obtained from the drone's FPV camera during flight. The real-time gesture recognition results displayed on the laptop interface during system operation are shown in Figure 9.

Figure 8. Captured Perimeter View

Figure 9. Real-time gesture detection output displayed on the laptop

Figure 10 presents field documentation of the hand gesture-based drone control process during outdoor testing. The sequence illustrates the initial condition of the drone before receiving any gesture command, followed by the Arm gesture used to activate the drone, and the Hover gesture demonstrating stable flight after successful initialization. This sequence represents the transition of the drone from an inactive state to an operational state under real field conditions. During the experiment, the distance between the operator and the drone was maintained at about 1-2 meters, allowing for reliable gesture detection and stable drone response in outdoor conditions. Based on the field testing results, the proposed hand gesture-based human–drone interaction system demonstrates reliable performance in outdoor agricultural environments. The system achieved an accuracy of 99.18%, with a real-time processing speed of 109 FPS and an average latency of 9.16 ms. These results indicate that the system is capable of responding quickly and consistently during perimeter monitoring tasks, even under varying outdoor lighting conditions. When compared with previous studies, the proposed system shows improved performance in terms of practical field deployment. A gesture recognition system based on CNN achieved an accuracy of around 92%, but required controlled backgrounds and stable lighting conditions. This performance is mainly achieved through the use of MediaPipe-based hand landmark extraction combined with an SVM classifier, which provides robust feature representation while maintaining low computational complexity for real-time processing.

Meanwhile, gesture-based control using Leap Motion sensors reported high success rates for certain static gestures, such as 100% for swiping left and right, while other gestures showed lower success rates, including 78.3% for swiping up and 93.3% for swiping down. However, these approaches rely on additional hardware and are generally designed for indoor environments. In contrast, the proposed system uses only a standard RGB camera, does not require a fixed background, and is able to operate in real outdoor conditions. This makes the system more suitable for agricultural applications, where lighting and environmental conditions are difficult to control. A quantitative comparison between the proposed system and previous studies is summarized in Table 4.

(a) (b) (c)

Figure 10. Field implementation of hand gesture-based drone control (a) drone condition before being controlled, (b) drone after executing the Arm gesture, and (c) drone maintaining a stable Hover position.

Table 4. Comparison with Previous Gesture-Based Control Systems

Study	Method	Enviroment	Additional Hardware	Accuracy /Succes Rate	Frame per Second (FPS) & Latency
M. Fuad et al. [9]	CNN	Indoor (controlled)	Yes	92%	Not reported
Tsai et al [11]	Leap Motion Based Sensor	Indoor	Yes	78.3% - 100%	Not reported

Preposed System	MediaPipe + Support Vector Machine (SVM)	Outdoor	No	99.18%	109 FPS / 9.16 ms

CONCLUSIONS AND SUGGESTIONS

This study developed a human–drone interaction system for perimeter monitoring in cornfields. The system uses hand gesture recognition based on MediaPipe Hands and the SVM algorithm. It operates in real time with low computational complexity and performs stably under outdoor lighting. Experimental results show an accuracy of 99.18%. The detection speed reaches 109 FPS, with an average latency of 9.16 ms. These results indicate that the system is responsive and suitable for real-time agricultural applications. Hand gesture-based drone control shows strong potential for supporting smart farming. It can improve field monitoring efficiency and reduce the need for manual supervision. The intuitive interaction removes the need for extra hardware. As a result, the proposed system offers practical and cost-effective agricultural monitoring. Still, several improvements remain for future work. The gesture set can be expanded for more complex drone operations. Gesture commands may also be limited to high-level instructions, combined with autonomous navigation using GPS-based or vision-based perimeter patrol. Additionally, implementation on embedded platforms like Raspberry Pi or Jetson Nano is recommended. This could enhance portability and autonomy for large-scale agricultural use.

REFERENCES

[1] F. A. Almalki, B. O. Soufiene, S. H. Alsamhi, and H. Sakli, "A low-cost platform for environmental smart farming monitoring system based on iot and uavs," *Sustainability (Switzerland)*, vol. 13, no. 11, Jun. 2021, doi: 10.3390/su13115908.

[2] A. Hafeez *et al.*, "Implementation of drone technology for farm monitoring & pesticide spraying: A review," Jun. 01, 2023, *China Agricultural University*. doi: 10.1016/j.inpa.2022.02.002.

[3] M. Yoo *et al.*, "Motion Estimation and Hand Gesture Recognition-Based Human–UAV Interaction Approach in Real Time," *Sensors*, vol. 22, no. 7, Apr. 2022, doi: 10.3390/s22072513.

[4] G. Ipate, C. Tudora, and F. Ilie, "Digital Analysis with the Help of an Integrated UAV System for the Surveillance of Fruit and Wine Areas," *Agriculture (Switzerland)*, vol. 14, no. 11, Nov. 2024, doi: 10.3390/agriculture14111930.

[5] P. Srinil and P. Thongnim, "Deep Learning Enhanced Hand Gesture Recognition for Efficient Drone use in Agriculture," 2024. [Online]. Available: www.ijacsa.thesai.org

[6] M. Anggraeni, H. Andhika F. R., H. Rante, and S. Sukaridhoto, "Indonesian Sign Language (SIBI) Learning Media Application Based on Deep Learning Technology for Deaf Children," *Journal of Technology and Informatics (JoTI)*, vol. 5, no. 1, pp. 41–47, Oct. 2023, doi: 10.37802/joti.v5i1.384.

[7] B. Taylor, M. Allen, P. Henson, X. Gao, H. Malik, and P. Zhu, "Enhancing Drone Navigation and Control: Gesture-Based Piloting, Obstacle Avoidance, and 3D Trajectory Mapping," *Applied Sciences (Switzerland)*, vol. 15, no. 13, Jul. 2025, doi: 10.3390/app15137340.

[8] N. Chalista Imanuela Natun, M. Angelica Santhia, and Y. R. Kaesmetan, "Identifikasi Pengenalan Wajah Berdasarkan Jenis Kelamin Menggunakan Metode Convolutional Neural Network (CNN)," *Journal of Technology and Informatics (JoTI)*, vol. 6, no. 1, pp. 50–57, Oct. 2024, doi: 10.37802/joti.v6i1.694.

[9] M. Fuad *et al.*, "Towards Controlling Mobile Robot Using Upper Human Body Gesture

Based on Convolutional Neural Network," *Journal of Robotics and Control (JRC)*, vol. 4, no. 6, pp. 856–867, 2023, doi: 10.18196/jrc.v4i6.20399.

[10] S. Abdalla and S. Baidya, "UAV Control with Vision-based Hand Gesture Recognition over Edge-Computing," May 2025, [Online]. Available: http://arxiv.org/abs/2505.17303

[11] *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2020.

[12] B. Hu and J. Wang, "Deep Learning Based Hand Gesture Recognition and UAV Flight Controls," *International Journal of Automation and Computing*, vol. 17, no. 1, pp. 17–29, Feb. 2020, doi: 10.1007/s11633-019-1194-7.

[13] J. W. Lee and K. H. Yu, "Wearable Drone Controller: Machine Learning-Based Hand Gesture Recognition and Vibrotactile Feedback," *Sensors*, vol. 23, no. 5, Mar. 2023, doi: 10.3390/s23052666.

[14] D. Tezza and M. Andujar, "The State-of-the-Art of Human-Drone Interaction: A Survey," *IEEE Access*, vol. 7, pp. 167438–167454, 2019, doi: 10.1109/ACCESS.2019.2953900.

[15] A. Azka, A. Santoso, and T. Agustinah, "Controlling a Quadcopter with Static Loads and Dynamic Wind Disturbances using a Fuzzy Controller," 2024.

[16] M. S. Abdallah, G. H. Samaan, A. R. Wadie, F. Makhmudov, and Y. I. Cho, "Light-Weight Deep Learning Techniques with Advanced Processing for Real-Time Hand Gesture Recognition," *Sensors*, vol. 23, no. 1, Jan. 2023, doi: 10.3390/s23010002.

[17] A. Edet, S. Inyang, I. Umoren, and U. E. Etuk, "Machine Learning Approach for Classification of Cyber Threats Actors in Web Region," *Journal of Technology and Informatics (JoTI)*, vol. 6, no. 1, pp. 70–77, Oct. 2024, doi: 10.37802/joti.v6i1.679.

[18] F. Zhang *et al.*, "MediaPipe Hands: On-device Real-time Hand Tracking," Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.10214

[19] "Hand landmarks detection guide | Google AI Edge | Google AI for Developers." Accessed: Oct. 30, 2025. [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker

[20] Lukman Arif Sanjani, R. Bimo Mandala Putra, and U. Laili Yuhana, "Exploring the Application of Machine Learning for Automatic Inbound Email Classification in CRM System at XYZ Company," *Journal of Technology and Informatics (JoTI)*, vol. 6, no. 1, pp. 1–7, Oct. 2024, doi: 10.37802/joti.v6i1.715.