Expert Systems With Applications 289 (2025) 128176

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Enhancing firefighter safety and efficiency through UAV-assisted AI-based human motion recognition system

Hong Wang [a], Chenyang Zhao [b], Yuan Feng [c], Xu Huang [c], Chengyi Qu [d], Yaguang Zhu [b], Ming Xin [c], Wenbin Guo [e,*]

[a] School of Artificial Intelligence, Hubei University, Wuhan 430062, China
[b] School of Construction Machinery, Chang'an University, Xi'an 710064, China
[c] Department of Mechanical and Aerospace Engineering, University of Missouri, Columbia, MO 65211, USA
[d] U.A. Whitaker College of Engineering, Florida Gulf Coast University, Florida 33965, USA
[e] College of Medicine, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

Keywords:
AI
Firefighter safety
Human motion recognition
UAVs

ABSTRACT

Fires pose grave threats to lives, properties, and natural resources, necessitating effective fire suppression and rescue efforts. This paper proposes a solution to enhance firefighter safety and efficiency by integrating Unmanned Aerial Vehicles (UAVs) with Artificial Intelligence (AI)-based human motion recognition. The proposed system employs supervised machine learning algorithms trained on firefighting video datasets to recognize six crucial firefighter actions: 'advance', 'climb', 'connect', 'fetchwater', 'hoselaying', and 'watersupply'. The trained model demonstrates 74.8% accuracy in identifying these actions during fire suppression activities. By utilizing UAVs and AI, this system offers extended detection range, deployment flexibility, and live, high-resolution image capture without risking pilots' lives. The integration of UAVs and AI holds great promise in mitigating the devastating impact of fires and safeguarding the lives of firefighters.

1. Introduction

Fires pose significant threats to lives, properties, and natural resources, necessitating effective fire suppression and rescue efforts (Barmpoutis, Papaioannou, et al., 2020). In the United States alone, the National Fire Protection Association (NFPA) reported over 1.39 million fires attended by local fire departments in 2023, resulting in 3,670 fatalities, 13,350 injuries, and property damage costs $23 billion (Hall, 2024). As the first responders to fire incidents, firefighters face perilous conditions when combating fires, with close proximity to fire sources (Morris & Chander, 2018; Page et al., 2019). Therefore, it is paramount to enhance firefighters' functionality and ensure their safety by recognizing their action patterns.

Early fire detection and continuous monitoring play pivotal roles in reducing casualties and supporting firefighting efforts (Chen et al., 2022; Fonollosa et al., 2018; Sousa et al., 2020; Sungheetha & Sharma R, 2020). Various technologies, including ground-based systems, manned aircraft, satellite-based systems, and UAVs have been employed for fire detection and monitoring (Mazzeo et al., 2022; Sudhakar et al., 2020). However, each of these systems comes with its limitations. Ground monitoring systems have constrained surveillance ranges due to terrain restrictions, while human-piloted aircraft require experienced personnel and are exposed to hazardous fire environments (Barmpoutis, Stathaki, et al., 2020; Crawford & Shahroudi, 2018). Satellite-based systems, although capable of covering vast areas, suffer from limited image resolution and frequency (Thangavel et al., 2023). In recent years, UAVs have become valuable tools for enhancing firefighter performance by tracking fire behavior and predicting its progression (Bailon-Ruiz et al., 2022; Ghali et al., 2022). UAVs provide numerous benefits, such as greater deployment flexibility, cost efficiency, extended detection range, elimination of pilot risk, and the capability to capture real-time, high-resolution images.

Despite growing interest in utilizing UAVs and AI to enhance firefighter safety and working efficiency, there is a lack of focus on real-time recognition of individual firefighters' actions and precise locations, particularly within the context of firefighting operations. Effective communication and coordination are crucial for firefighters due to the hazardous and rapidly changing conditions of their work environments. Nevertheless, effective coordination remains a challenge and warrants further exploration. Firefighters acquire essential skills related to

* Corresponding author.
E-mail address: wenbin.guo@ufl.edu (W. Guo).

https://doi.org/10.1016/j.eswa.2025.128176
Received 24 March 2025; Received in revised form 2 May 2025; Accepted 12 May 2025
Available online 16 May 2025

information gathering, coordination optimization, and sensemaking predominantly through their experience (Aust et al., 2023; Wolbers et al., 2018). This gap limits the ability to provide crucial insight and support and fully leverage the potential to improve firefighters' effectiveness and safeguard their lives.

To bolster firefighter safety and working efficiency, the integration of UAVs and AI-based human motion detection has shown great promise (Seraj & Gombolay, 2020). This study proposes the utilization of UAVs equipped with advanced AI-based algorithms to recognize firefighters' action patterns during fire suppression activities. By utilizing existing human motion datasets from firefighting videos, we successfully trained a supervised machine learning algorithm to recognize six key firefighter actions: 'advance', 'climb', 'connect', 'fetchwater', 'hoselaying', and 'watersupply'. The trained model exhibited high accuracy when tested on existing video footage. Integrating this model into a UAV system could greatly enhance firefighter efficiency and safety during fire response operations.

This study aims to develop and evaluate a UAV-assisted AI system for recognizing firefighter actions to enhance safety and efficiency. This paper outlines the methodology and findings of the UAV-assisted AI-based human motion recognition system (Fig. 1), highlighting its potential applications in enhancing firefighter safety and working efficiency. In the realm of firefighting operations, a fully autonomous inspection and analysis system based on UAV is essential for improving situational awareness and response efficiency. The system operates through a meticulously orchestrated five-step process. It begins with data acquisition, where the UAV employs sensors and cameras to gather real-time information from firefighters. Then, the acquired data undergoes comprehensive analysis, utilizing advanced algorithms to identify key patterns and potential hazards. The system proceeds to the classification phase, where it categorizes the analyzed information, distinguishing firefighter actions. Once the classification is complete, the UAV receives instructions based on identified risks and optimal response strategies. Finally, the current system integrates data acquisition, analysis, classification, and instruction to support real-time action recognition and decision-making. UAV locomotion, envisioned as a future extension, would enable the platform to autonomously navigate firefighting sites and provide dynamic, situational support to ground personnel. However, the implementation and validation of such autonomous navigation would require further system integration, multi-agent coordination, and field-level testing, which are beyond the scope of the current study. This modular and extensible architecture is designed to enhance firefighting teams' situational awareness and enable a more proactive approach to risk mitigation in future deployments. By harnessing the power of UAVs and AI, our goal is to support ongoing efforts to reduce the destructive effects of fires and safeguard the lives of those who bravely fight them.

This paper makes the following contributions. First, this paper proposes a novel UAV-assisted, AI-based human motion recognition system to improve firefighter safety and operational efficiency. The system integrates UAVs with AI-based motion recognition to address limitations in existing fire monitoring tools, particularly the lack of real-time individual action recognition. It features offline model training, real-time inference on a Jetson Nano onboard the UAV, and autonomous responses, marking a first-of-its-kind integration of AI in an aerial system for firefighting support.

Second, this paper demonstrates the effectiveness of AI models in classifying key firefighter actions. Supervised learning algorithms, including traditional classifiers and a deep neural network (DNN), were trained on firefighter video datasets. The DNN achieved the highest accuracy of 74.8% across six critical actions: 'advance', 'climb', 'connect', 'fetchwater', 'hoselaying', and 'watersupply'. Real-time testing using Jetson Nano and a Raspberry Pi camera confirmed the system's ability to identify actions accurately in both recorded and live settings.

Third, this paper provides empirical insights into system performance and practical challenges. The study compares machine learning model accuracies, presents confusion matrices, and highlights common misclassifications, especially among similar actions. This study also reveals challenges such as partial visibility, overlapping movements, and limitations from dataset size and diversity.

The rest of the paper is organized as follows: Section 2 reviews the relevant literature on UAV and AI applications in firefighting, highlighting the gap addressed by this research; Section 3 describes the proposed system design and construction, detailing the machine learning algorithms used, the UAV hardware, and the software platform; Section 4 presents the results and discussion, evaluating the performance of the machine learning models and the UAV's real-time recognition capabilities; and Section 5 concludes the paper with a summary of findings and potential future work.

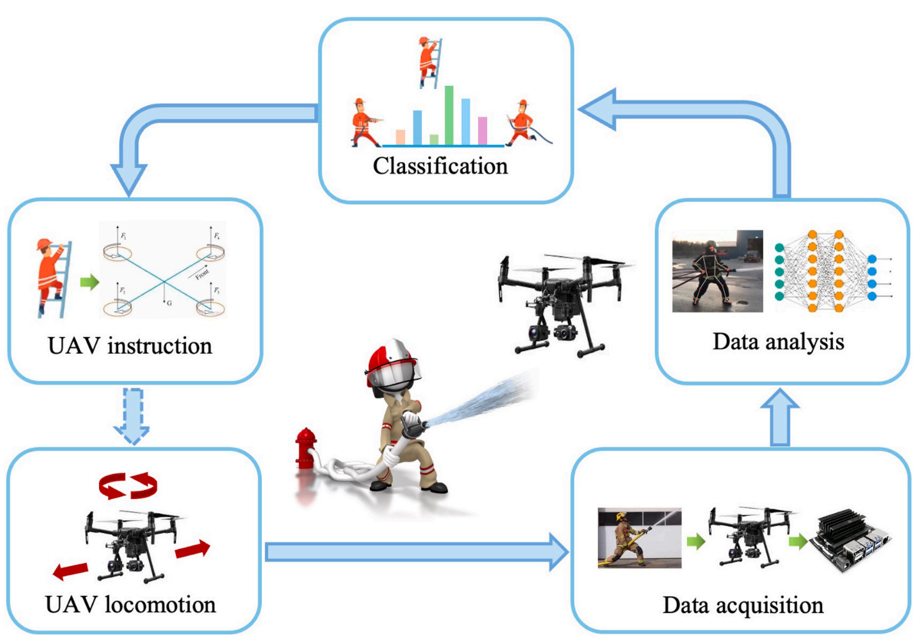

Fig. 1. A fully autonomous inspection and analysis system based on UAV for firefighters' postures.

2. Related works

2.1. AI and UAV applications in firefighting response

AI can be used to enhance the capabilities of UAVs, making them more autonomous, efficient, and effective (Puente-Castro et al., 2022; Rezwan & Choi, 2022). One important application of AI in the UAV field is autonomous navigation. By using machine learning algorithms, UAVs are capable of autonomously navigating through complex environments, avoiding obstacles, and optimizing their flight paths. Advances in AI for object detection and identification have enabled UAVs to be used in tasks such as search and rescue. In (Afridi et al., 2019), UAVs were used for rescue operations in floods and tsunamis by detecting and tracking humans and water. Despite these advances, there is a clear research gap in AI-UAV systems that can interpret human motion during dynamic and hazardous missions. Current platforms focus on environmental cues rather than analyzing the activity of ground personnel to adapt aerial behavior accordingly.

UAVs have been employed for various tasks across multiple sectors, including emergency response. They provide a range of advantages for firefighting operations. In the realm of reconnaissance and surveillance, UAVs provide images with higher resolution and shorter time intervals compared to satellites, thereby improving the response time of firefighters (Roldán-Gómez et al., 2021; Yuan et al., 2015). In terms of direct fire suppression, UAVs can be equipped with water or other fire retardants and deployed in areas that are difficult or dangerous for human firefighters to reach (Cervantes et al., 2018; Peña et al., 2022; Zadeh et al., 2021). In post-fire assessments, UAVs equipped with various sensors can help evaluate the extent of fire damage by capturing high-resolution images and videos (Carvajal-Ramírez et al., 2019). However, while these systems enhance spatial coverage and environmental awareness, they typically rely on manual or semi-autonomous operations, lacking intelligent, human-centric decision-making capabilities that could adapt dynamically to firefighter status or actions.

AI is a rapidly developing field that is beginning to impact all aspects of people's lives. One important application of AI in the fire protection field is predictive analytics and fire forecasting (Lee et al., 2025). By using historical data from fire modeling, AI can provide potential algorithms for detecting and forecasting fire scenarios (Huang et al., 2022). AI can be used in the fire command scene to improve on-site command (Chang et al., 2022). Recent advances in deep learning have also demonstrated the capacity of AI models, such as artificial neural networks (ANN), convolutional neural networks (CNN), and long short-term memory (LSTM) networks, to predict climate-related variables that can influence fire risks, including temperature, precipitation, and air quality (Guo et al., 2023, 2025; Guo, He, & Wang, 2024; Guo, He, Wang, et al., 2024; He & Guo, 2024). These capabilities strengthen fire prevention efforts by offering timely and accurate predictions. Another notable domain is firefighter training: AI-enhanced simulations replicate real-world emergencies, enabling safe and effective scenario practice. These systems can adapt to individual physical data to personalize training intensity and strategies (Xu et al., 2020). Nonetheless, these AI applications have primarily focused on environmental or system-level predictions, with limited direct integration with human-centric data, particularly firefighter motion during operations.

2.2. AI in human motion recognition and situational awareness

Human motion recognition aims to understand and classify human behaviors, and can be implemented based on various data modalities, which are mainly divided into visual modalities (such as RGB, skeleton, depth, point cloud, event stream, etc.) and non-visual modalities (such as audio, acceleration, WiFi, radar, etc.). Among them, RGB images and skeleton data have become the most mainstream research direction due to their complementary characteristics: RGB images provide rich appearance and background information, which are suitable for scene understanding; skeleton data extracts human joint structures through pose estimation, possessing advantages in noise resistance and structural representation, making it suitable for efficient and fast human behavior modeling. In the RGB modality, the two-stream convolutional network proposed by Simonyan and Zisserman uses RGB frames and optical flow to model appearance and motion features respectively, and is considered a foundational work (Simonyan & Zisserman, 2014). Karpathy et al. further reduced computational cost by introducing multi-resolution cropping strategies (Karpathy et al., 2014).

In terms of spatiotemporal modeling, 3D Convolutional Neural Networks (3D CNNs) are used to directly extract spatiotemporal features from videos. The C3D model proposed by Tran et al. was the first to validate the effectiveness of 3D convolution in action recognition (Tran et al., 2015), but its high computational cost became a bottleneck. Subsequent work, such as that by Varol et al., proposed to expand the receptive field in the temporal dimension to enhance modeling capability (Varol et al., 2018), while Qiu et al. decomposed 3D convolution into 2D spatial convolution and 1D temporal convolution to improve efficiency (Qiu et al., 2017). In addition, some studies have transformed human joints into pseudo-images or video clips, which are then input into 2D or 3D CNNs for modeling (Choutas et al., 2018; Yan et al., 2019). Among them, the PoseC3D method has shown excellent performance on several skeleton-based action recognition benchmarks, demonstrating strong robustness and interoperability, although its 3D CNN backbone is relatively heavy (Duan et al., 2022).

The importance of maintaining direct and prompt communication and coordination among emergency personnel, particularly in life-threatening situations, cannot be overstated (Mistick et al., 2022; Stanton et al., 2001). Improving situational awareness and learning from past incidents can significantly boost the safety of firefighters, a key aspect often emphasized in their training regimen (Page & Butler, 2018). In (Page & Butler, 2018), the location of fatal burnover was studied with the finding that the geospatial environmental data can help firefighters to locate fatal locations. Situational awareness refers to the ability to recognize, process, and understand the key information regarding the team's status in relation to the mission (Endsley, 1995). Recent research has been conducted on enhancing the safety of firefighters by deepening the understanding of various facets of situational awareness. In (Jolly & Freeborn, 2017), fire radiative power variations measured by satellite are linked to the fire behavior risk rating prediction to keep firefighters safe. UAVs can improve firefighters' operational capabilities by overcoming the physical limitations of human responders, allowing them to be deployed in various tasks. One of the main applications of UAVs is surveillance on site, detecting and observing forest fires (Reusen et al., 2008). In Perez-Saura et al. (Perez-Saura et al., 2023), a solution was offered for launching fire suppression devices via windows from a UAV in urban firefighting. Building on related advancements in human movement analysis (Wang et al., 2024), AI models, originally developed for emotion recognition through movement in performance settings, can also be adapted to interpret firefighter posture. To enhance situational awareness, UAV gathered information can be integrated into ERIS (Emergency Response Information System) (Agrawal et al., 2020). Despite the current research emphasis on aiding firefighters with improved situational awareness, there is a lack of focus on real-time recognition of individual firefighters' actions and precise locations.

2.3. Human-machine collaboration in AI-driven UAV systems

Human-machine collaboration in UAV systems is advancing toward more adaptive and user-centered designs to support complex, multi-agent missions. Lim et al. (2021) and Feuerriegel et al. (2021) emphasize adaptive interaction frameworks and interface design as keys to enabling one operator to manage multiple UAVs efficiently. Their findings highlight the importance of context-aware autonomy, reduced cognitive load, and interfaces that promote situational awareness and trust. Hoc (2001) contributes a cognitive framework that stresses shared mental models and mutual predictability, essential for effective cooperation in dynamic environments.

Building on these foundations, Cao et al. (2024) propose enhanced control strategies for multi-UAV systems, incorporating dynamic task allocation and feedback loops to support flexible human oversight. Zhou and Liu (2021) demonstrate long-term collaboration through persistent person tracking, showing how adaptive systems can learn from human input to maintain performance in real-world conditions. Wang et al. (2023) introduced an AI-based action detection system mounted on UAVs, designed to monitor firefighters in real time and identify hazardous behaviors. Together, these studies reflect a shift toward collaborative autonomy, where UAVs and humans operate as coordinated teammates, with system design focused on adaptability, transparency, and cognitive compatibility.

In summary, our solution represents a shift from traditional environment-focused UAV deployments to human-centric awareness enhancement, bridging a critical gap in current firefighting technologies. As summarized in Table 1, our system uniquely incorporates real-time motion recognition of firefighters using onboard AI processing, an aspect absent in existing UAV and AI integration efforts. Compared to prior studies that focus on either environmental monitoring or predictive analytics, our work contributes a novel, end-to-end approach that emphasizes immediate operational awareness and decision support tailored to individual firefighter actions.

3. System design and construction

The proposed system enhances firefighting operations through the integration of four key modules, as illustrated in Fig. 2: (1) Model Module, (2) Camera Module, (3) Jetson Nano Module, and (4) STM32 Control Module. Together, these components form a cohesive pipeline for gesture-based UAV control in real time.

Model Module: This module outlines the offline training phase conducted on a local computer. Machine learning models are trained to recognize and interpret firefighter gestures and movement patterns using a video dataset. The resulting model captures complex motion features relevant to firefighting contexts. Once trained, the model and its associated parameters are transferred to the Jetson Nano device for deployment.

Camera Module: A miniature onboard camera provides real-time video input to the Jetson Nano. This camera captures continuous visual data from the environment, enabling the system to monitor firefighter activities. The live video feed is streamed directly into the Jetson Nano for immediate processing.

Jetson Nano Module: Acting as the computational core of the system, the Jetson Nano hosts the trained recognition model. It processes incoming video data in real time to extract motion features, predict the firefighter's action category, and calculate corresponding key coordinates. The Jetson Nano's compact form factor and computational capability make it well-suited for onboard deployment on UAVs, enabling low-latency decision-making in dynamic environments.

STM32 Control Module: This module comprises the STM32 microcontroller and the UAV's control logic. It receives the action category and key coordinate data from the Jetson Nano. Based on this information, the module generates scenario-specific control commands, allowing the UAV to autonomously respond to firefighter gestures. This real-time closed-loop system empowers the UAV to support firefighting personnel with agility and precision.

This design emphasizes the synergy between offline learning and real-time deployment, enabling a practical and responsive tool for field operations. Note that detailed UAV posture and flight mechanics are outside the scope of this work and are planned for future research in context-specific scenarios.

3.1. *Machine learning algorithm*

3.1.1. *Data preparation*

Similar to most AI-based methods, the proposed approach is divided into three main stages: data preparation, training and testing, and automated annotation, all designed to create a reliable system for detecting changes in firefighter actions within video sequences. In the initial stage of data preparation, diverse and representative datasets are collected and preprocessed to ensure optimal model training. We utilized a dataset of firefighter operational scenes collected from the internet. The dataset comprises six categories of firefighter actions: 'advance', 'climb', 'connect', 'fetchwater', 'hoselaying', and 'watersupply'. Prior to data augmentation, the dataset contained a total of 128 video samples, with 106 used for training and the remainder for testing. Each video lasted approximately 5s and was recorded at 30 frames per second (FPS).

To address class imbalance and enhance model generalization, data augmentation was applied. As shown in Table 2, each action category originally included 10 to 34 training videos, yielding a total of 13,353 training frames and 3,889 testing frames. Following augmentation, each category was expanded to 60–204 training videos and 12 testing videos, with the number of testing frames increased to approximately 2,000 per class.

Each original video was augmented approximately sixfold through mirroring and scaling transformations (at factors of 0.8× and 1.2×), which are well-suited for skeleton-based data. Other augmentation strategies, such as temporal cropping, were deliberately excluded to preserve the structural completeness of the skeleton and ensure the reliability of keypoint extraction.

In total, the augmented dataset consists of 636 training videos and 72 testing videos, corresponding to 92,555 training frames and 12,017 testing frames. These enhancements ensure that each action category is represented by a comparable number of samples and frames, thereby mitigating the effects of data imbalance. The updated statistics are summarized in Table 2 below.

Table 1
Comparative analysis of key related works and our contribution.

Study / system	UAV function	AI component	Integration focus	Real-time processing	Human motion recognition	Novelty compared to prior work
Roldán-Gómez et al. (2021)	Recon & surveillance	None	Spatial data acquisition	No	No	High-res imagery for mapping
Cervantes et al. (2018)	Fire suppression	None	Payload deployment	No	No	UAV in hard-to-reach areas
Lee et al. (2025)	None	Fire forecasting	Environmental modeling	No	No	Predictive analytics
Afridi et al. (2019)	Search & rescue	Object detection	Victim tracking	Partial	No	AI for human detection in SAR
Agrawal et al. (2020)	Surveillance & mapping	Fire behavior prediction	ERIS integration	Partial	No	AI-driven risk mapping
This Study	Recon, tracking, safety	Human action recognition	Firefighter-centric data loop	**Yes**	**Yes**	**First real-time UAV-based firefighter action detection system with on-board AI**

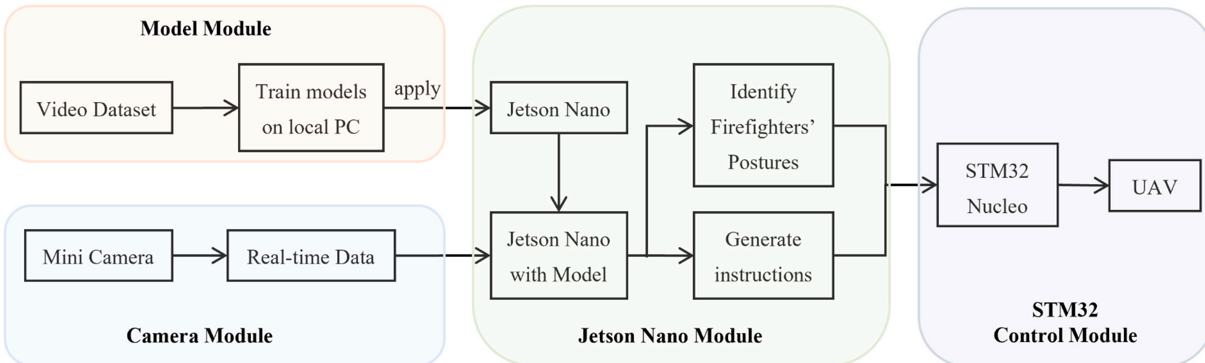

Fig. 2. UAV-assisted AI-based Human Motion Recognition System.

Table 2
Summary of training and testing video data before and after data augmentation.

Behavior	Train Vid (Pre)	Train Frm (Pre)	Test Vid (Pre)	Test Frm (Pre)	Train Vid (Post)	Train Frm (Post)	Test Vid (Post)	Test Frm (Post)
Advance	10	1444	2	347	60	8440	12	1818
Climb	10	1156	2	318	60	7033	12	1714
Connect	10	1403	2	352	60	8087	12	2018
Fetchwater	28	4187	6	1119	168	25,022	12	2232
Hoselaying	34	4775	6	1021	204	28,821	12	2045
Watersupply	14	2388	4	732	84	14,152	12	2190

Abbreviations: Vid = number of videos, Frm = number of frames, Pre = before augmentation, Post = after augmentation.

3.1.2. Feature extraction

Based on the augmented dataset, skeletal keypoint data are used to perform frame-level behavior classification. For each frame, 33 human body keypoints are extracted using MediaPipe, with the x and y coordinates of each keypoint forming a 66-dimensional static feature vector.

To extract dynamic features, keypoints are sampled at 10-frame intervals to form temporal data groups. Within each group, four inter-joint distances are computed: left shoulder–right shoulder (LS–RS), left shoulder–left elbow (LS–LE), left elbow–left wrist (LE–LW), and left hip–left knee (LH–LK). The distance with the smallest standard deviation across the group is selected as the normalization reference.

Subsequently, six additional inter-joint distances are computed: left ankle–left hip (LA–LH), right ankle–right hip (RA–RH), left wrist–nose (LW–N), right wrist–nose (RW–N), left wrist–left hip (LW–LH), and right wrist–right hip (RW–RH). In addition, the velocities and accelerations of six keypoints (LA, RA, LH, RH, LW, RW) are calculated over time. After normalization based on the reference distance and standard human body proportions, each group yields 18 dynamic features.

For feature fusion, the average dynamic feature vector (per video) is concatenated with each frame's static feature vector, resulting in a combined 84-dimensional feature vector per frame.

3.1.3. Model implementation

The training and testing phases involve the use of five traditional machine learning models: Naïve Bayes, Support Vector Machines (SVM), Logistic Regression, Decision Trees, and Random Forest. Naïve Bayes is a probabilistic classifier based on Bayes' theorem, while SVM aims to find the optimal hyperplane that separates different classes in feature space. Logistic Regression predicts the probability of a binary outcome, Decision Trees make decisions based on feature values, and Random Forest is an ensemble of decision trees that improves accuracy. Additionally, a DNN, which is a feed-forward neural network with input, hidden, and output layers, is trained for more complex pattern recognition. Table 3 compares the features for traditional machine learning models and DNN architecture.

The flowchart in Fig. 3 illustrates the deep learning action recognition system, which encapsulates the process from training videos to action detection. The input to the network consists of 84-dimensional fused features, which incorporate both static and dynamic information extracted from each video frame. The DNN architecture comprises four hidden layers with 128, 64, 32, and 16 neurons, respectively, each using the ReLU activation function, followed by an output layer of 6 neurons activated by the Softmax function to classify the six defined firefighter actions. Given that feature extraction has already been performed through prior processing, the neural network serves a focused role in feature abstraction and classification, rather than low-level feature learning (Fig. 4).

The process begins with the preprocessed training data, where the network learns high-level motion representations from the input

Table 3
Features comparison for traditional machine learning models and DNN architecture.

Aspect	Naïve bayes	SVM	Logistic regression	Decision trees	Random forest	DNN
Model type	Probabilistic classifier	Margin-based classifier	Linear probabilistic model	Tree-based	Ensemble of decision trees	Feed-forward neural network
Learning objective	Bayes theorem	Maximize margin between classes	Estimate class probabilities	Recursive partitioning	Majority vote across trees	Learn complex feature hierarchies
Interpretability	High	Moderate	High	High	Moderate	Low
Training time	Fast	Medium to slow	Fast	Fast	Medium to slow	Slow
Scalability	High	Moderate	High	Moderate	High	Moderate to High
Best for	Simple distributions	High-dimensional separable data	Binary/multiclass with linear trends	Easily interpretable trees	Handling noise & variance	Complex, non-linear motion recognition

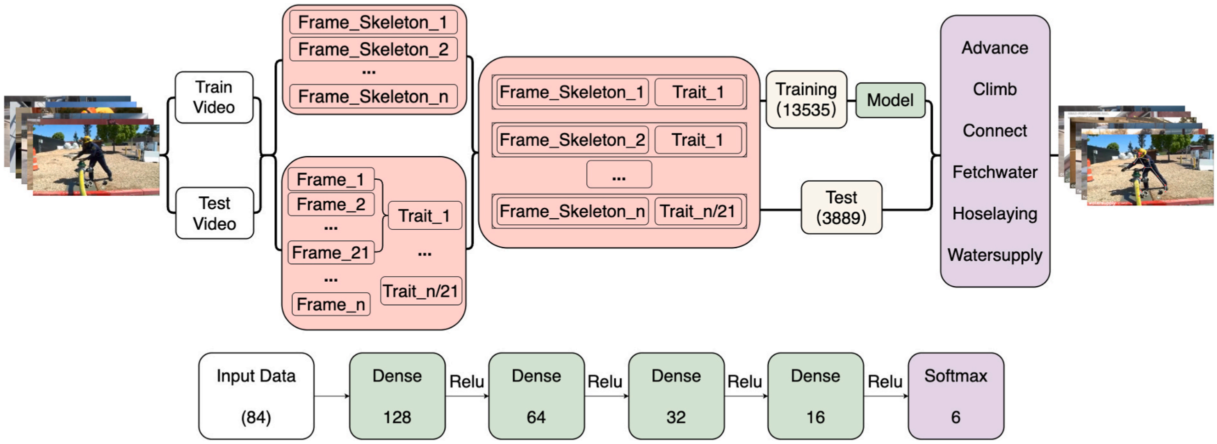

Fig. 3. Flowchart of deep learning action recognition system.

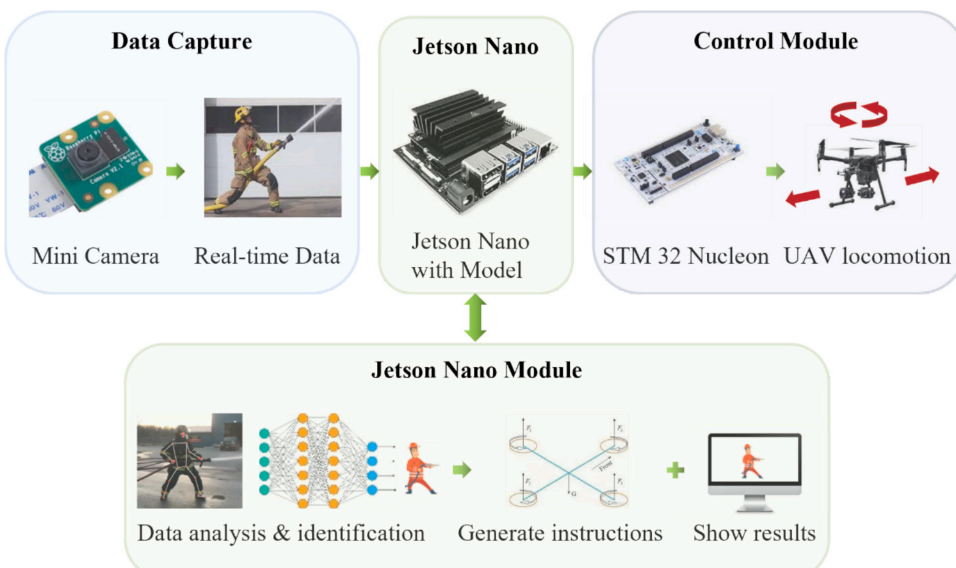

Fig. 4. UAV-assisted AI-based Human Motion Recognition System.

features. Each layer in the network contributes to progressively refining the internal representation of the action patterns. Once trained, the network is applied to test videos, where it performs action classification in real time by mapping the input features to the most probable action class.

This structured pipeline, comprising data input, layered feature transformation, and final output generation, demonstrates the effectiveness of a lightweight, fully connected DNN architecture in recognizing complex motion patterns within real-world UAV video scenarios.

3.1.4. Evaluation metrics

The performance of the developed system is evaluated using key multiclass classification metrics, including precision, recall, accuracy, and F1 score. In the multiclass setting, precision quantifies the system's ability to correctly predict instances of each class, with an emphasis on reducing false positives across all classes. Recall assesses the system's sensitivity by measuring its effectiveness in identifying all true instances of each class, thereby minimizing false negatives. Both metrics are calculated on a per-class basis and then averaged using macro approaches. The F1 score, defined as the harmonic mean of precision and recall, provides a balanced measure of the model's performance, especially useful in cases where class distribution is imbalanced. Accuracy, on the other hand, reflects the overall proportion of correctly classified samples across all classes. All metrics range from 0 to 1, where higher values indicate stronger performance. Together, these metrics offer a comprehensive assessment of the system's ability to recognize and classify actions across multiple categories. The corresponding formulas are presented below:

$$\text{Precision}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} \tag{1}$$

$$\text{Recall}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \tag{2}$$

$$\text{F1-score}_i = 2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{3}$$

where:

TP_i : True Positives for class i

FP_i : False Positives for class i

FN_i : False Negatives for class i

Then, depending on the averaging method:

- Macro-averaged Precision = Average of per-class precisions
- Macro-averaged Recall = Average of per-class recalls
- Macro-averaged F1-score = Average of per-class F1 scores

Thus, in multiclass, accuracy is usually better thought of simply as:

$$\text{Accuracy} = \frac{\text{number of correct predictions}}{\text{total number of predictions}} = \frac{\sum \text{TP}_i}{\text{Total}} \tag{4}$$

3.2. UAV hardware

In this research, a UAV was built to obtain real-time image data and respond simultaneously to the identification results. The UAV primarily consists of two parts: computing modules and mechanical components.

3.2.1. Computing module

The UAV computing module contains three parts, i.e., Nvidia Jetson Nano, Raspberry Pi Camera, and STM32 Nucleo board. All these devices are embedded into the UAV as a payload. A detailed description of each module is given as follows:

(i) Nvidia Jetson Nano:

A microcomputer with 128 parallel processing cores, which is enough to support real-time video feed (Cass, 2020). It boasts a heterogeneous CPU–GPU architecture, offering a compact form factor, lightweight design, and low power consumption. The Jetson Nano is among the most commonly used edge devices, equipped with accelerators for machine learning inference (Assunção et al., 2022). It is also commonly used for various neural network applications, including object detection, image segmentation and classification (Gomes et al., 2022; Salih & Basman Gh., 2020). In comparison, the popular Raspberry Pi 4 has the same 4 GB of RAM as the Jetson Nano and a better CPU, but it has limitations when running AI algorithms. For example, in Ayoub's work (Ayoub & Schneider-Kamp, 2021), they compared Raspberry Pi 4 and Jetson Nano to run the Yolov4 model and pointed out that Raspberry Pi 4 cannot infer normally due to excessive memory requirements. Jetson Nano, by contrast, boasts a powerful GPU with significant AI advantages, making it well-suited for deep learning tasks. Its practicality and suitability for AI and machine learning projects are evident. This component plays a critical role in real-time firefighter detection, data analysis, and motion recognition. It efficiently translates classification results into UAV commands, facilitating seamless communication with the UAV's control module.

(ii) Raspberry Pi Camera v2:

The compact camera boasts an impressive Sony IMX219 8-megapixel sensor, supporting the 1080p30 video mode through the Raspberry Pi Camera Module 2 (Pagnutti et al., 2017). Operating as a CSI camera, it utilizes specialized data channels and instructions dedicated to image acquisition and processing, setting it apart from USB cameras. This specialized design results in superior performance and accelerated transfer rates, meeting the demands of high frame rates and resolutions. Leveraging host visual acceleration hardware, it significantly reduces CPU usage. When integrated with the Jetson Nano, this camera seamlessly provides real-time video data for precise analysis and identification.

(iii) STM32 Nucleo board:

In order to receive instructions and control the actions of the UAV, there is a need to have STM32 Nucleo board embedded on UAV to achieve instruction transmission. STM32 has a mainstream cortex core, extremely high performance, rich peripherals and software packages, and an ARM core. Compared with the 51 microcontroller, it has a higher working frequency, stronger computing power, and richer resources.

3.2.2. Mechanical components

In Fig. 5, the construction of the drone is shown. The drone frame, crafted from the Tarot 650 Sport Quadcopter Build Kit, incorporates a configured Tarot gimbal below and positions a high-capacity 4S lithium-ion polymer battery at the gimbal's rear. Autonomous flight requires the drone to be equipped with essential computing hardware. Hence, the frame's imperative traits entail sturdiness and lightness, attributes facilitated by the application of carbon fiber, rendering this Tarot frame under 500 g in weight.

Atop this frame resides the DJI N3 flight controller kit, comprising a main control unit, a GPS-compass unit, and a battery management unit. Further above the N3, a customized PCB houses a buck module and an STM32 socket. The buck module serves to reduce the battery voltage to 5 V, providing power to both the Jetson Nano and the S.BUS remote control receiver.

3.3. Software platform

The well-trained firefighters' posture classification model was trained with machine learning algorithms, as mentioned in Section 3.1, on the PC and then deployed on the Jetson Nano. With the help of Raspberry Pi v2 camera and Jetson Nano's high performance running machine learning models and rapid processing of video images, real-time monitoring of firefighters' postures is achieved. Some important features are calculated every 10 frames, such as speed, acceleration, etc., to better identify firefighter postures. The extraction and use of dynamic features, including speed and acceleration, have been described in detail in the earlier Feature Extraction section. The model can detect firefighters shown in the camera feed and extract skeletons of the firefighter to predict the class of the current posture. The identification results, including the skeleton, bounding box, predicted class, and accuracy, are displayed. Meanwhile, the Jetson Nano generates a set of flight control instructions, based on the identification of the posture. These targeted instructions essentially assign values to the different directional axes of the drone to achieve a complete attitude response. STM32 will then synthesize the received signals into UAV execution instructions and send

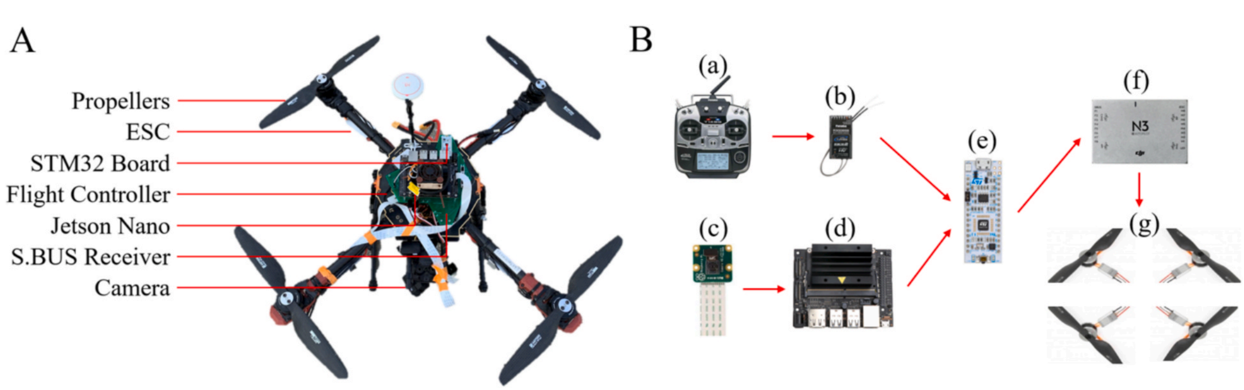

Fig. 5. Mechanical system and task sequence. (A) Hardware components. (B) The system architecture: (a) S.BUS remote controller (b) S.BUS signal receiver (c) Raspberry Pi camera module 2 (d) Nvidia Jetson Nano (e) STM32 (f) DJI N3 flight controller (g) Propeller, Motor and ESCs.

them to the UAV flight control module to complete the response.

4. Results and discussion

4.1. Machine learning simulation

Table 4 shows the comparative analysis of these models reveals that the DNN outperforms the others, achieving the highest accuracy at 74.8%. This performance stands in contrast to the accuracies obtained by Naïve Bayes (39.1%), SVM (51.1%), Logistic Regression (49.7%), Decision Tree (42.7%), and Random Forest (52.6%). The findings indicate that DNN is particularly effective in extracting and analyzing high-level, concealed, and significant features in the context of firefighter action classification. For detailed information on the hyperparameters of the five traditional machine learning models, please refer to Supplementary Information Table S1. And for the performance of the five traditional machine learning models using 5-fold cross-validation, please refer to Table S2.

Fig. 6 illustrates a comparison of six machine learning algorithms, namely Naïve Bayes, SVM, logistic regression, decision tree, random forest, and DNN. In examination of the confusion matrix the DNN demonstrates relatively accurate predictions across six distinct patterns. However, accurately predicting true labels 0, 1, and 2 proves challenging due to their movement patterns being similar to the other three activities. These labels not only have the highest rate of false predictions but also contribute to an imbalanced matrix in the DNN model. By contrast, models such as Logistic Regression, SVM, Random Forest, and Decision Tree exhibited more balanced performance in the confusion matrix specific to firefighter action detection. Further analysis reveals that false predictions are often associated with frames containing multiple firefighters. Common issues included limited visibility of firefighters' bodies, with parts such as feet or arms being obscured by equipment, fire trucks, or other firefighters. These obstructions often resulted in zero-value data in movement analysis, posing a challenge for feature extraction. For the per-class precision, recall, and F1 score of the five traditional machine learning models, please refer to Supplementary Information Table S3.

Fig. 7 presents illustrative examples of the results derived from a single frame in one of the input videos. The image showcases the automatically detected human skeleton along with corresponding key points overlaid on the scene. These six image annotations encompass important actions such as 'advance', 'climb', 'connect', 'fetchwater', 'hoselaying', and 'watersupply'. Through the developed algorithms, firefighter actions are successfully annotated for every frame in the input video, with the generated labels accurately reflecting the specific actions taking place in each frame.

4.2. UAV real-time recognition

Our test data include a set of six distinct postures frequently adopted by firefighters in the field: 'advance', 'climb', 'watersupply', 'fetchwater', 'hoselaying', and 'connect'. The efficacy of the system was gauged through an analytical assessment of both recorded videos of firefighters performing these motions and direct live demonstrations. Our assessment procedure is encapsulated in Fig. 8 (a) to (f), which delineate the recognition results obtained from video data. The empirical data demonstrate that the system is able to consistently identify five out of the six postures with good accuracy, with the actions 'advance' and 'fetchwater' being recognized with heightened precision.

Table 4
Performance comparison of various models for machine learning methods.

Algorithms	Precision	Recall	Accuracy	F1 score
Naive Bayes	43.6 %	38.9 %	39.1 %	37.8 %
SVM	47.7 %	48.6 %	51.1 %	44.7 %
Logistic Regression	47.3 %	47.9 %	49.7 %	45.3 %
Decision Tree	41.6 %	41.4 %	42.7 %	40.2 %
Random Forest	70.7 %	51.6 %	52.6 %	51.7 %
DNN	73.2 %	73.0 %	74.8 %	72.4 %

For the live demonstration experiments, the setup, procedure, and visualization outputs are described as follows. During the UAV-based real-time detection trials, the UAV was positioned at a height of 1.2 m above ground level, maintaining a horizontal orientation without tilt and facing directly toward the test subject. The onboard camera, stabilized via a gimbal, provided a steady, level, forward-facing view. The test subjects stood approximately 3 m in front of the UAV to ensure that their entire body remained fully visible within the captured frame.Each subject conducted five complete testing rounds. In each round, they performed six firefighter-specific actions in different sequences, with each action repeated 3 times from various directions. This setup ensured that the detection cycle captured complete action processes, provided sufficient testing samples for each action to assess model accuracy, and minimized disruptions caused by non-standard actions or transitions.

As shown in the experimental result figures, skeletal keypoint detection, subject positioning, action recognition results, recognition accuracy, and the probability scores for each action were displayed in real time during the detection process. To better document the results and facilitate subsequent analysis, the detection interface was screen-recorded, capturing both the video feed and real-time recognition outputs. Recorded videos were later reviewed to observe and analyze the prediction results for each tested action, with sample results illustrated in the provided figures.

To support real-time application, the system is implemented with a two-stage inference pipeline. During the data acquisition stage, 10 consecutive frames are accumulated (based on a 30 FPS stream), resulting in a theoretical minimum acquisition time of 333 ms. In the inference stage, action recognition is performed on the Jetson Nano (GPU mode), with an average inference time of 170 ms per cycle. This leads to an end-to-end latency of approximately 500 ms, corresponding to a real-time detection frequency of 2 Hz. The Jetson Nano operates with a unified memory usage of 3.0 GB out of 4 GB, maintaining a 75% utilization rate during operation.

In comparative analyses between recorded video data and live demonstrations, as depicted in Fig. 8 (g) to (l), the performance metrics for most action classes when tested in real-time conditions are consistent with those observed in video simulations. Notably, the actions 'advance' and 'hoselaying' demonstrate superior recognition accuracy. Conversely, the action 'climb' is frequently misclassified as 'hoselaying'. The classifications for 'watersupply' and 'fetchwater' exhibit lower accuracy in live testing compared to their respective recorded video-based results.

From the analysis of the two distinct sets of real-time test results, it is evident that the system successfully achieves its design objective by simultaneously recognizing firefighters' motions, though there remains substantial room for enhancement. Within the assessed motion categories, the classifications for 'advance' and 'hoselaying' exhibit robust performance, whereas the 'connect' and 'watersupply' classifications present challenges, often merging together in scenarios that necessitate precise differentiation, a task that proves to be complex. Furthermore, the composition and diversity of the dataset significantly influence the model's accuracy and overall performance efficacy. Due to the limited availability of publicly accessible video data, a compromise is necessary to balance data volume with class distribution. The frequent overlap and similarity of firefighters' actions, coupled with the absence of standardized movements for each category in practical settings, pose significant obstacles in constructing a well-defined categorical dataset for each action class. Consequently, this often leads to misclassifications, particularly within the 'fetchwater' category, as it is frequently confused with other motion classes.

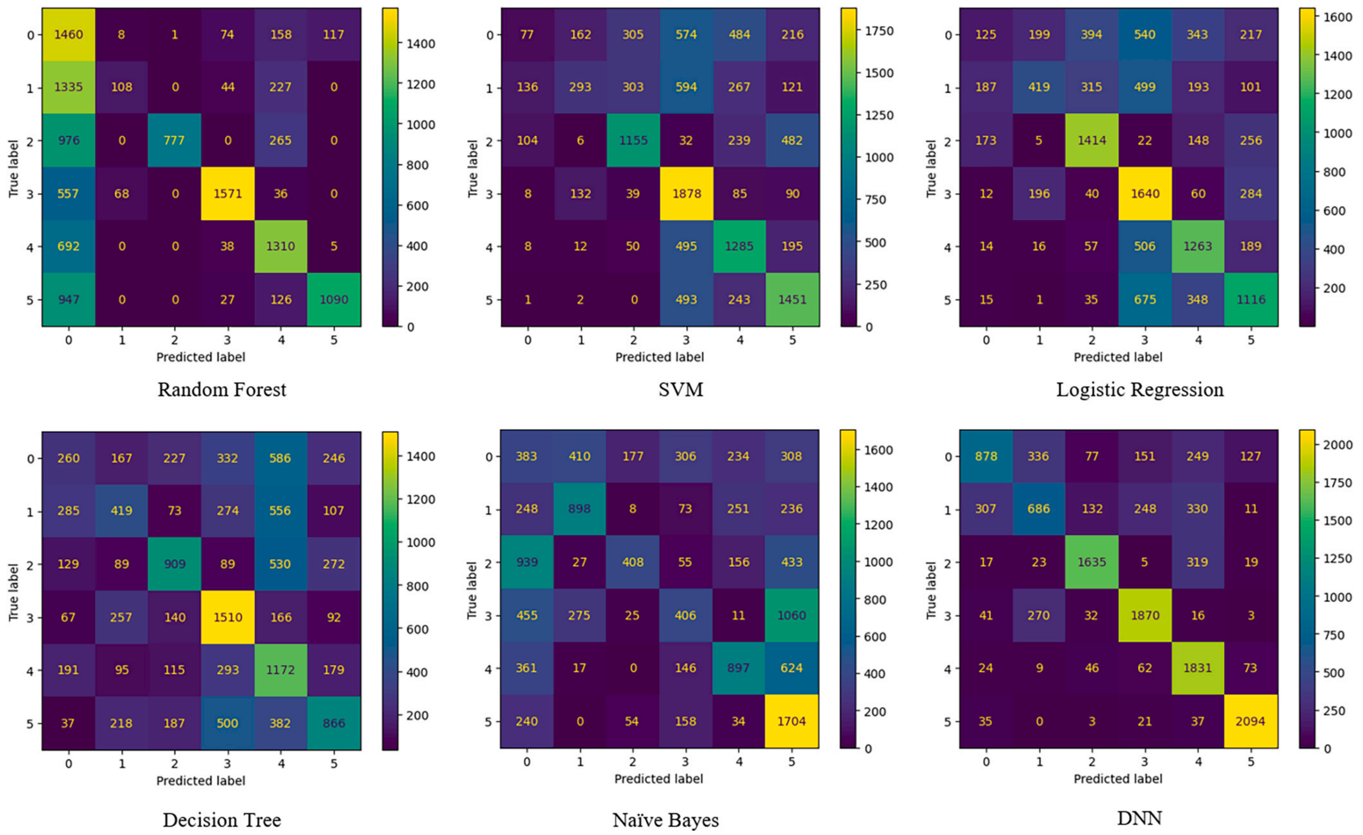

Fig. 6. Confusion Matrix of six machine learning algorithms, the numbers 0 to 5 represent the following actions, respectively: 'advance', 'climb', 'connect', 'fetchwater', 'hoselaying' and 'watersupply'.

Fig. 7. The automated annotation on videos: 'advance', 'climb', 'connect', 'fetchwater', 'hoselaying' and 'watersupply'.

5. Conclusion

In this study, a UAV and AI-based system is developed to detect and classify the operations of firefighters. The UAV is adept at discerning actions routinely executed by firefighters, facilitating understanding and support mechanisms that ensure their safety during firefighting endeavors and bolster their capability in analyzing the progression of fire incidents. This constitutes an inaugural integration of AI within an aerial hardware-software ensemble to support firefighting personnel. Our results show that the DNN yields the highest accuracy in firefighter action classification but faces challenges in predicting specific movements and handling multi-firefighter obstructions. While the system performs well in recognizing 'advance' and 'hoselaying,' it struggles with differentiating 'connect' and 'watersupply' due to overlapping actions and dataset limitations, emphasizing the need for further refinement. Looking forward, we anticipate the inclusion of multifaceted scenarios such as urban conflagrations, forest fires, and aquatic blaze responses within the ambit of this AI-augmented firefighting

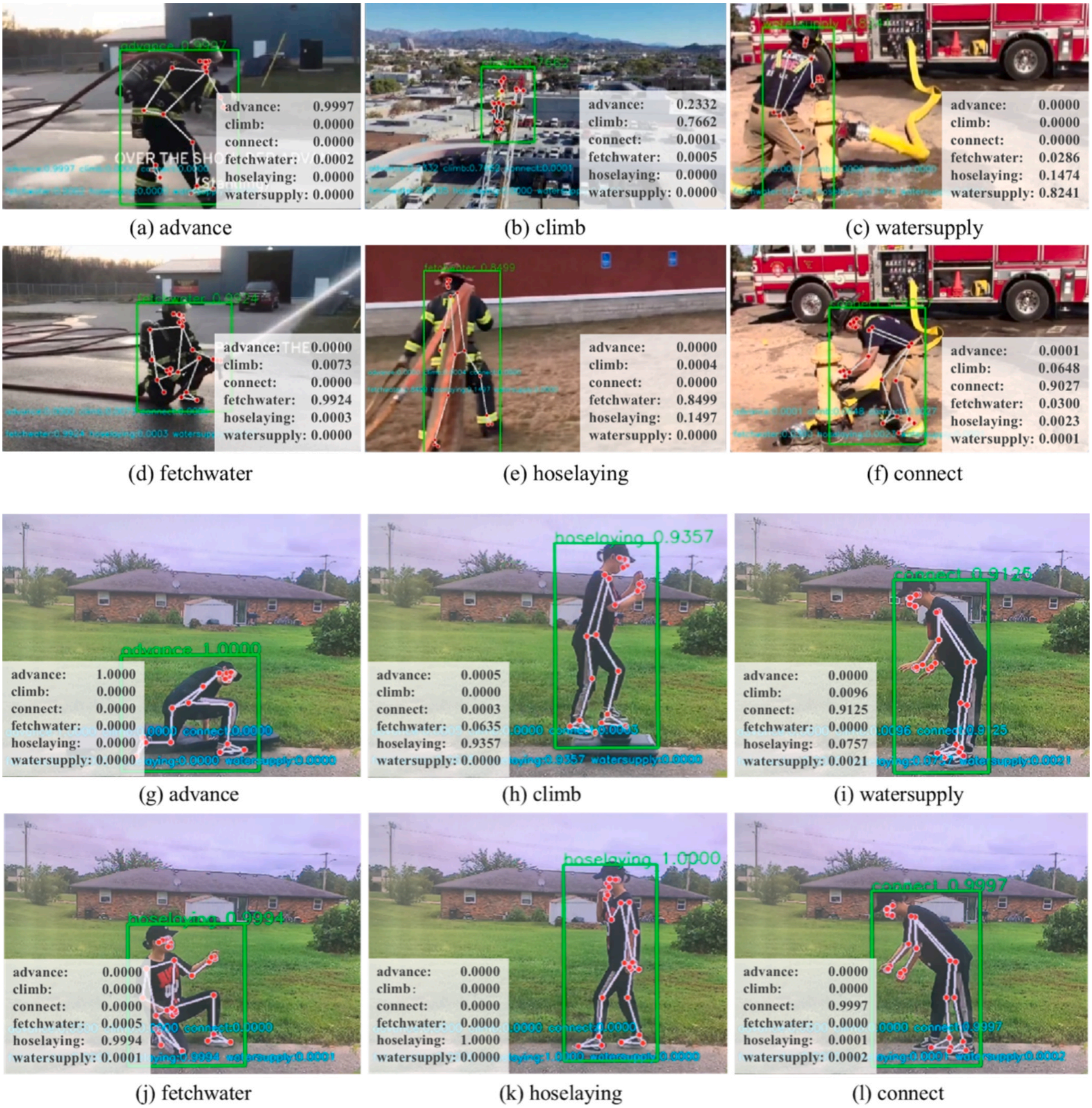

Fig. 8. Results from the UAV-assisted AI-based Human Motion Recognition System are presented, categorizing actions into six types: 'advance', 'climb', 'watersupply', 'fetchwater', 'hoselaying', and 'connect'. The dataset comprises twelve images, where (a) to (f) depict outcomes derived from recorded video data, showcasing the system's ability to accurately identify and classify actions. The (g) to (l) demonstrate the system's performance in real-time conditions for the same actions. This distinction highlights the system's robustness in both controlled and dynamic environments.

system. The aspiration is to evolve the UAV into a more potent and versatile entity capable of executing a diverse array of tasks, necessitating the development of more intricate, task-specific mechanical frameworks, alongside precise sensing and control apparatus, with the ultimate goal of saving lives and safeguarding property.

CRediT authorship contribution statement

Hong Wang: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing. **Chenyang Zhao:** Data curation, Formal analysis, Investigation, Methodology, Software, Validation. **Yuan Feng:** Data curation, Investigation, Methodology, Software, Validation, Visualization. **Xu Huang:** Data curation, Investigation, Writing. **Chengyi Qu:** Project administration, Supervision. **Yaguang Zhu:** Project administration, Supervision. **Ming Xin:** Writing – review & editing. **Wenbin Guo:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing.

Funding

This study was supported by the Natural Science Foundation for Young Scientists of Shannxi Province, China (Grant No. 2024JC-YBQN-0476).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2025.128176.

Data availability

Data will be made available on request.

References

Afridi, A., Minallah, N., Sami, I., Allah, M., Ali, Z., & Ullah, S. (2019). Flood Rescue Operations Using Artificially Intelligent UAVs. In *2019 15th International Conference on Emerging Technologies (ICET)* (pp. 1–5). https://doi.org/10.1109/ICET48972.2019.8994394

Agrawal, A., Abraham, S., Burger, B., Christine, C., Fraser, L., Hoeksema, J., Hwang, S., Travnik, E., Kumar, S., Scheirer, W., Cleland-Huang, J., Vierhauser, M., Bauer, R., & Cox, S. (2020). The Next Generation of Human-Drone Partnerships: Co-Designing an Emergency Response System. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). https://doi.org/10.1145/3313831.3376825

Assunção, E., Gaspar, P. D., Mesquita, R., Simões, M. P., Alibabaei, K., Veiros, A., & Proença, H. (2022). Real-time weed control application using a jetson nano edge device and a spray mechanism. *Remote Sensing, 14*(17), 4217. https://doi.org/10.3390/rs14174217

Aust, F., Heinemann, L., Holtz, M., Hagemann, V., & Peifer, C. (2023). Team flow among firefighters: Associations with collective orientation, teamwork-related stressors, and resources. *International Journal of Applied Positive Psychology, 8*(2), 339–363. https://doi.org/10.1007/s41042-023-00093-7

Ayoub, N., & Schneider-Kamp, P. (2021). Real-time on-board deep learning fault detection for autonomous UAV inspections. *Electronics, 10*(9), 1091. https://doi.org/10.3390/electronics10091091

Bailon-Ruiz, R., Bit-Monnot, A., & Lacroix, S. (2022). Real-time wildfire monitoring with a fleet of UAVs. *Robotics and Autonomous Systems, 152*, Article 104071. https://doi.org/10.1016/j.robot.2022.104071

Barmpoutis, P., Papaioannou, P., Dimitropoulos, K., & Grammalidis, N. (2020). A review on early forest fire detection systems using optical remote sensing. *Sensors, 20*(22), 6442. https://doi.org/10.3390/s20226442

Barmpoutis, P., Stathaki, T., Dimitropoulos, K., & Grammalidis, N. (2020). Early fire detection based on aerial 360-degree sensors, deep convolution neural networks and exploitation of fire dynamic textures. *Remote Sensing, 12*(19), 3177. https://doi.org/10.3390/rs12193177

Cao, Z., Zhang, B., Qiu, J., & Jin, R. (2024). Enhanced human-UAV collaborative control strategies for multi-UAVs in complex task environments. *IEEE International Conference on Unmanned Systems (ICUS), 2024*, 421–428. https://doi.org/10.1109/ICUS61736.2024.10839784

Carvajal-Ramírez, F., Marques Da Silva, J. R., Agüera-Vega, F., Martínez-Carricondo, P., Serrano, J., & Moral, F. J. (2019). Evaluation of fire severity indices based on pre- and post-fire multispectral imagery sensed from UAV. *Remote Sensing, 11*(9), 993. https://doi.org/10.3390/rs11090993

Cass, S. (2020). Nvidia makes it easy to embed AI: The Jetson nano packs a lot of machine-learning power into DIY projects - [Hands on]. *IEEE Spectrum, 57*(7), 14–16. https://doi.org/10.1109/MSPEC.2020.9126102

Cervantes, A., Garcia, P., Herrera, C., Morales, E., Tarriba, F., Tena, E., & Ponce, H. (2018). A Conceptual Design of a Firefighter Drone. In *2018 15th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)* (pp. 1–5). https://doi.org/10.1109/ICEEE.2018.8533926

Chang, R. H., Peng, Y.-T., Choi, S., & Cai, C. (2022). Applying Artificial Intelligence (AI) to improve fire response activities. *Emergency Management Science and Technology, 2* (1), 1–6. https://doi.org/10.48130/EMST-2022-0007

Chen, X., Hopkins, B., Wang, H., O'Neill, L., Afghah, F., Razi, A., Fulé, P., Coen, J., Rowell, E., & Watts, A. (2022). Wildland fire detection and monitoring using a drone-collected RGB/IR image dataset. *IEEE Access, 10*, 121301–121317. https://doi.org/10.1109/ACCESS.2022.3222805

Choutas, V., Weinzaepfel, P., Revaud, J., & Schmid, C. (2018, June). PoTion: Pose MoTion Representation for Action Recognition. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT. doi: 10.1109/cvpr.2018.00734.

Crawford, S. W., & Shahroudi, K. E. (2018). Wildfire detection and communication–aerospace applications–trade study. *Insight, 21*(4), 32–40. https://doi.org/10.1002/inst.12224

Duan, H., Zhao, Y., Chen, K., Lin, D., & Dai, B. (2022). Revisiting skeleton-based action recognition. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022*, 2959–2968. https://doi.org/10.1109/cvpr52688.2022.00298

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Feuerriegel, S., Geraldes, R., Goncalves, A., Liu, Z., & Prendinger, H. (2021). Interface design for human-machine collaborations in drone management. *IEEE Access, 9*, 107462–107475. https://doi.org/10.1109/ACCESS.2021.3100712

Fonollosa, J., Solórzano, A., & Marco, S. (2018). Chemical sensor systems and associated algorithms for fire detection: A review. *Sensors, 18*(2), 553. https://doi.org/10.3390/s18020553

Ghali, R., Akhloufi, M. A., & Mseddi, W. S. (2022). Deep learning and transformer approaches for UAV-based wildfire detection and segmentation. *Sensors, 22*(5), 1977. https://doi.org/10.3390/s22051977

Gomes, H., Redinha, N., Lavado, N., & Mendes, M. (2022). Counting people and bicycles in real time using YOLO on Jetson nano. *Energies, 15*(23), 8816. https://doi.org/10.3390/en15238816

Guo, Q., He, Z., & Wang, Z. (2023). Prediction of monthly average and extreme atmospheric temperatures in Zhengzhou based on artificial neural network and deep learning models. *Frontiers in Forests and Global Change, 6*. https://doi.org/10.3389/ffgc.2023.1249300

Guo, Q., He, Z., & Wang, Z. (2024). Monthly climate prediction using deep convolutional neural network and long short-term memory. *Scientific Reports, 14*(1). https://doi.org/10.1038/s41598-024-68906-6

Guo, Q., He, Z., & Wang, Z. (2025). Assessing the effectiveness of long short-term memory and artificial neural network in predicting daily ozone concentrations in Liaocheng City. *Scientific Reports, 15*(1). https://doi.org/10.1038/s41598-025-91329-w

Guo, Q., He, Z., Wang, Z., Qiao, S., Zhu, J., & Chen, J. (2024). A performance comparison study on climate prediction in Weifang city using different deep learning models. *Water, 16*(19), 2870. https://doi.org/10.3390/w16192870

Hall, S. (2024, October 31). Fire loss in the United States. *National Fire Protection Association (NFPA)*. https://www.nfpa.org/education-and-research/research/nfpa-research/fire-statistical-reports/fire-loss-in-the-united-states.

He, Z., & Guo, Q. (2024). Comparative analysis of multiple deep learning models for forecasting monthly ambient PM2.5 concentrations: A case study in Dezhou City, China. *Atmosphere, 15*(12), 1432. https://doi.org/10.3390/atmos15121432

Hoc, J.-M. (2001). Towards a cognitive approach to human–machine cooperation in dynamic situations. *International Journal of Human-Computer Studies, 54*(4), 509–540. https://doi.org/10.1006/ijhc.2000.0454

Huang, X., Wu, X., & Usmani, A. (2022). Perspectives of Using Artificial Intelligence in Building Fire Safety. In M. Naser, & G. Corbett (Eds.), *Handbook of Cognitive and Autonomous Systems for Fire Resilient Infrastructures* (pp. 139–159). Springer International Publishing. https://doi.org/10.1007/978-3-030-98685-8_6.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014, June). Large-Scale Video Classification with Convolutional Neural Networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition.* 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA. doi: 10.1109/cvpr.2014.223.

Jolly, W. M., & Freeborn, P. H. (2017). Towards improving wildland firefighter situational awareness through daily fire behaviour risk assessments in the US Northern Rockies and Northern Great Basin. *International Journal of Wildland Fire, 26* (7), 574. https://doi.org/10.1071/WF16153

Lee, D., Perez Tellez, F., & Jaiswal, R. (2025). Predicting fire incidents with ML: An XAI approach. *AI and Ethics*. https://doi.org/10.1007/s43681-025-00683-y

Lim, Y., Pongsakornsathien, N., Gardi, A., Sabatini, R., Kistan, T., Ezer, N., & Bursch, D. J. (2021). Adaptive human-robot interactions for multiple unmanned aerial vehicles. *Robotics, 10*(1), 12. https://doi.org/10.3390/robotics10010012

Mazzeo, G., De Santis, F., Falconieri, A., Filizzola, C., Lacava, T., Lanorte, A., Marchese, F., Nolè, G., Pergola, N., Pietrapertosa, C., & Satriano, V. (2022). Integrated satellite system for fire detection and prioritization. *Remote Sensing, 14*(2), 335. https://doi.org/10.3390/rs14020335

Mistick, K. A., Dennison, P. E., Campbell, M. J., & Thompson, M. P. (2022). Using geographic information to analyze wildland firefighter situational awareness: impacts of spatial resolution on visibility assessment. *Fire, 5*(5), 151. https://doi.org/10.3390/fire5050151

Morris, C. E., & Chander, H. (2018). The impact of firefighter physical fitness on job performance: A review of the factors that influence fire suppression safety and success. *Safety, 4*(4), 60. https://doi.org/10.3390/safety4040060

Page, W. G., & Butler, B. W. (2018). Fuel and topographic influences on wildland firefighter burnover fatalities in Southern California. *International Journal of Wildland Fire, 27*(3), 141. https://doi.org/10.1071/WF17147

Page, W. G., Freeborn, P. H., Butler, B. W., & Jolly, W. M. (2019). A review of US wildland firefighter entrapments: Trends, important environmental factors and research needs. *International Journal of Wildland Fire, 28*(8), 551. https://doi.org/10.1071/WF19022

Pagnutti, M., Ryan, R. E., Cazenavette, G., Gold, M., Harlan, R., Leggett, E., & Pagnutti, J. (2017). Laying the foundation to use Raspberry Pi 3 V2 camera module imagery for scientific and engineering purposes. *Journal of Electronic Imaging, 26*(1), Article 013014. https://doi.org/10.1117/1.JEI.26.1.013014

Peña, P. F., Ragab, A. R., Luna, M. A., Ale Isaac, M. S., & Campoy, P. (2022). WILD HOPPER: A heavy-duty UAV for day and night firefighting operations. *Heliyon, 8*(6), Article e09588. https://doi.org/10.1016/j.heliyon.2022.e09588

Perez-Saura, D., Fernandez-Cortizas, M., Perez-Segui, R., Arias-Perez, P., & Campoy, P. (2023). Urban firefighting drones: Precise throwing from UAV. *Journal of Intelligent & Robotic Systems, 108*(4), 66. https://doi.org/10.1007/s10846-023-01883-6

Puente-Castro, A., Rivero, D., Pazos, A., & Fernandez-Blanco, E. (2022). A review of artificial intelligence applied to path planning in UAV swarms. *Neural Computing and Applications, 34*(1), 153–170. https://doi.org/10.1007/s00521-021-06569-4

Qiu, Z., Yao, T., & Mei, T. (2017). *Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks* (No. arXiv:1711.10305). arXiv. doi: 10.48550/arXiv.1711.10305.

Reusen, I., Lewyckyj, N., Adriaensen, S., Biesemans, J., Everaerts, J., & Kempenaers, S. (2008). *Near-real-time forest fires monitoring system: Case study with a manned aerial vehicle within the OSIRIS project*. 145–152. doi: 10.2495/FIVA080151.

Rezwan, S., & Choi, W. (2022). Artificial intelligence approaches for UAV navigation: Recent advances and future challenges. *IEEE Access, 10*, 26320–26339. https://doi.org/10.1109/ACCESS.2022.3157626

Roldán-Gómez, J. J., González-Gironda, E., & Barrientos, A. (2021). A survey on robotic technologies for forest firefighting: Applying drone swarms to improve firefighters' efficiency and safety. *Applied Sciences, 11*(1), 363. https://doi.org/10.3390/app11010363

Salih, T. A., & Basman Gh, M. (2020). A novel face recognition system based on Jetson Nano developer kit. *IOP Conference Series: Materials Science and Engineering, 928*(3), Article 032051. https://doi.org/10.1088/1757-899X/928/3/032051

Seraj, E., & Gombolay, M. (2020). *Coordinated Control of UAVs for Human-Centered Active Sensing of Wildfires* (arXiv:2006.07969). arXiv. doi: 10.48550/arXiv.2006.07969.

Simonyan, K., & Zisserman, A. (2014). *Two-Stream Convolutional Networks for Action Recognition in Videos* (No. arXiv:1406.2199). arXiv. doi: 10.48550/arXiv.1406.2199.

Sousa, M. J., Moutinho, A., & Almeida, M. (2020). Thermal infrared sensing for near real-time data-driven fire detection and monitoring systems. *Sensors, 20*(23), 6803. https://doi.org/10.3390/s20236803

Stanton, N. A., Chambers, P. R. G., & Piggott, J. (2001). Situational awareness and safety. *Safety Science, 39*(3), 189–204. https://doi.org/10.1016/S0925-7535(01)00010-8

Sudhakar, S., Vijayakumar, V., Sathiya Kumar, C., Priya, V., Ravi, L., & Subramaniyaswamy, V. (2020). Unmanned aerial vehicle (UAV) based forest fire detection and monitoring for reducing false alarms in forest-fires. *Computer Communications, 149*, 1–16. https://doi.org/10.1016/j.comcom.2019.10.007

Sungheetha, A., & Sharma, R. R. (2020). Real time monitoring and fire detection using internet of things and cloud based drones. *Journal of Soft Computing Paradigm, 2*(3), 168–174. https://doi.org/10.36548/jscp.2020.3.004

Thangavel, K., Spiller, D., Sabatini, R., Amici, S., Sasidharan, S. T., Fayek, H., & Marzocca, P. (2023). Autonomous satellite wildfire detection using hyperspectral imagery and neural networks: A case study on Australian Wildfire. *Remote Sensing, 15*(3), 720. https://doi.org/10.3390/rs15030720

Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning spatiotemporal features with 3D convolutional networks. *IEEE International Conference on Computer Vision (ICCV), 2015*, 4489–4497. https://doi.org/10.1109/iccv.2015.510

Varol, G., Laptev, I., & Schmid, C. (2018). Long-term temporal convolutions for action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(6), 1510–1517. https://doi.org/10.1109/tpami.2017.2712608

Wang, H., Feng, Y., Huang, X., & Guo, W. (2023). An AI-Based Action Detection UAV System to Improve Firefighter Safety. In V. G. Duffy (Ed.), *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management* (Vol. 14028, pp. 632–641). Springer Nature Switzerland. doi: 10.1007/978-3-031-35741-1_44.

Wang, H., Zhao, C., Huang, X., Zhu, Y., Qu, C., & Guo, W. (2024). Emotion Recognition in Dance: A Novel Approach Using Laban Movement Analysis and Artificial Intelligence. In V. G. Duffy (Ed.), *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management* (Vol. 14710, pp. 189–201). Springer Nature Switzerland. doi: 10.1007/978-3-031-61063-9_13.

Wolbers, J., Boersma, K., & Groenewegen, P. (2018). Introducing a fragmentation perspective on coordination in crisis management. *Organization Studies, 39*(11), 1521–1546. https://doi.org/10.1177/0170840617717095

Xu, D., Song, Y., Meng, Y., István, B., & Gu, Y. (2020). Relationship between firefighter physical fitness and special ability performance: Predictive research based on machine learning algorithms. *International Journal of Environmental Research and Public Health, 17*(20), 7689. https://doi.org/10.3390/ijerph17207689

Yan, A., Wang, Y., Li, Z., & Qiao, Y. (2019). PA3D: Pose-action 3D machine for video recognition. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019*, 7914–7923. https://doi.org/10.1109/cvpr.2019.00811

Yuan, C., Zhang, Y., & Liu, Z. (2015). A survey on technologies for automatic forest fire monitoring, detection, and fighting using unmanned aerial vehicles and remote sensing techniques. *Canadian Journal of Forest Research, 45*(7), 783–792. https://doi.org/10.1139/cjfr-2014-0347

Zadeh, N. R. N., Abdulwakil, A. H., Amar, M. J. R., Durante, B., & Santos, C. V. N. R. (2021). Fire-fighting UAV with shooting mechanism of fire extinguishing ball for smart city. *Indonesian Journal of Electrical Engineering and Computer Science, 22*(3), 1320. https://doi.org/10.11591/ijeecs.v22.i3.pp1320-1326

Zhou, T., & Liu, Y. (2021). Long-term person tracking for unmanned aerial vehicle based on human-machine collaboration. *IEEE Access, 9*, 161181–161193. https://doi.org/10.1109/ACCESS.2021.3132077