



OPEN Real-time construction safety monitoring using a drone based deep hybrid attention model

K. Sujit¹, K. Indhumathi², G. Mohan³, N. N. Jose⁴, Kama Ramudu⁵, Tarek Salem Abdennaji⁶, N. Rao Cheepurupalli⁷, Taoufik Saidana⁸ & Yewuhalashet Fissaha^{7,9}

Maintaining worker safety on hazardous construction sites remains a serious challenge, especially due to the rampant shortage or misuse of Personal Protective Equipment (PPE), which is a major contributor to accidents in at-height operations. Conventional monitoring systems are usually not as adaptable or accurate as needed in real-world situations. To address this, a Discrete Dilated Cosine Causal Convolution Krawtchouk Orangutan Multi-Tchebichef Head Self-Attention Network (2D-3CKO-MTHSAN) is proposed, which incorporates Discrete Dilated Cosine Causal Convolution with Multi-Head Self-Attention (DCCMSA) and Discrete Cosine-Krawtchouk-Tchebichef Transform (DCKTKT) to enhance PPE detection accuracy and robustness. The approach uses real-time drone surveillance to continuously acquire visual data at actual construction sites and controlled lab settings, with emphasis on identifying the salient elements of the Personal Fall Arrest System (PFAS), such as helmets, harnesses, and lifelines. A two-stage preprocessing pipeline comprising entropy filtering and τ -Kendall correlation analysis improves image quality and feature prominence. Deep spatio-temporal features are extracted using Adaptive Causal Decision Transformers, and the DCCMSA-DCKTKT model parameters are adjusted using the Orangutan Optimization Algorithm (OOA) to achieve stable performance under changing environmental conditions. Experimental results confirm the system's superiority over conventional deep learning models, achieving 99.9% detection accuracy. Compared to other fixed CCTV image-based or single-stage CNN-based PPE detection methods with limited scalability, the new hybrid 2D-3CKO-MTHSAN approach integrates adaptive optimization and multi-scale attention in an innovative manner to enable real-time, accurate drone-based detection. This is an example of our solution's novelty in integrating discrete transforms with the Orangutan Optimization Algorithm (OOA) for efficient field deployment.

Keywords Worker safety in high-risk construction, Personal protective equipment, Adaptive causal decision transformers, Orangutan optimization algorithm, Discrete dilated cosine causal convolution

Construction is inherently hazardous, especially when working with heavy equipment, wobbly structures, and at heights. Most construction fatalities and injuries are directly caused by the occasional or improper use of Personal Protective Equipment (PPE), including lifelines, harnesses, and helmets. This is one of the most pressing safety concerns¹⁻³. The dynamic, high-speed conditions of today's building sites are beyond the capabilities of traditional safety observation systems, which are either manual or fixed-based CCTV installations. The need for more intelligent, automated monitoring systems is growing urgently because traditional approaches are not flexible enough for real-time monitoring and frequently fail to identify safety infractions in a timely manner.

¹Department of Electrical and Electronics Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India. ²Department of Computer Applications, Kalasalingam Academy of Research and Education, Krishnankoil, India. ³Department of Mathematics, K.S.Rangasamy College of Technology, Tiruchengode, Namakkal District, Tamilnadu, India. ⁴Denken Solutions, Clrvine, USA. ⁵Department of Electronics and Communication Engineering, Aditya University, Surampalem, Kakinada, Andhra Pradesh, India. ⁶Department of Civil Engineering, College of Engineering, Northern Border University, Arar, Saudi Arabia. ⁷Faculty of Mines, Aksum Institute of Technology, Aksum University, Tigray, Ethiopia. ⁸Center for Scientific Research and Entrepreneurship, Northern Border University, 73213 Arar, Saudi Arabia. ⁹Department of Electrical and Computer Engineering, National Institute of Technology, Asahikawa College, 2-2-1-6 Syunkodai, Asahikawa City, Hokkaido 071-8142, Japan. ✉email: sujitvj.kumar@gmail.com; nraocheepurupalli@gmail.com; yowaggaye@gmail.com

Due to aging workforces, extended fabrication times, and diminishing populations, the construction industry faces global productivity challenges^{4–6}.

Canada, Hong Kong, India, the US, and the EU are among the nations that suffer from low productivity. But with the advent of digital fabrication and robotics, productivity has almost doubled. Construction robots are classified into Cartesian (gantry) and articulated (robotic arm) systems based on their kinematics. While articulated systems can move beyond these dimensions, Cartesian systems travel along normal axes^{7–9}. Benefits of Cartesian systems include enormous build capacity, low cost, and ease of setup. Though expensive and complicated, articulated systems can increase filament adhesion. Multi-robot and mobile systems are required for real-scale structures made of various materials to overcome these constraints. Drones with deep learning capabilities are transforming safety on construction sites by offering mobility, flexibility, and a wide field of view^{10–12}. They can identify employees, evaluate PPE compliance, and highlight safety hazards when paired with advanced neural networks. Computer vision, image processing, and sophisticated algorithms are used in real-time drone-assisted safety monitoring to identify and assess on-site safety compliance. This technology is revolutionizing worker safety in the construction sector by lowering accidents and guaranteeing regulatory compliance^{13–15}.

Because the use of Personal Fall Arrest Systems (PFAS) is not consistently monitored, accidents — especially falls from heights — remain the leading cause of fatalities on construction sites, despite the introduction of stringent safety rules. Conventional safety inspection techniques are frequently labor-intensive, manual, and prone to human error, making it challenging to ensure real-time compliance with safety protocols. There is a critical need for an intelligent, automated system capable of accurately detecting safety equipment violations under diverse environmental conditions to enhance worker protection and reduce fatal incidents.

Research gap and motivation

Previous studies on PPE detection have mainly relied on conventional CNN-based or single-attention models that perform well in static or laboratory conditions but degrade under drone-captured imagery with motion blur, varying altitudes, and complex lighting. Furthermore, most existing frameworks detect only a limited number of PPE classes and lack adaptability for real-time on-site deployment due to high computational cost. Hence, there is a clear need for an integrated, lightweight model that can accurately recognize PPE and PFAS across multiple classes directly from aerial footage in dynamic construction environments.

Novelty and contribution

The novelty and contribution of this work are given below:

- A novel Discrete Dilated Cosine Causal Convolution Krawtchouk Orangutan Multi-Tchebichef Head Self-Attention Network (2D-3CKO-MTHSAN) framework is proposed, which integrates Discrete Dilated Cosine Causal Convolution with Multi-Head Self-Attention (DCCMSA) and Discrete Cosine-Krawtchouk-Tchebichef Transform (DCKTKT) to enable robust and accurate detection of safety equipment.
- The approach uses real-time drone surveillance to continuously acquire visual data at both construction sites and controlled laboratory settings, focusing on identifying Lifelines, harnesses, and helmets, which are essential components of the Personal Fall Arrest System (PFAS).
- Introduced a dual-step preprocessing technique combining entropy-based filtering and τ -Kendall correlation analysis to enhance the clarity and relevance of drone-captured visual data.
- Employed Adaptive causal decision transformers to extract deep spatio-temporal features from preprocessed construction site imagery.
- Applied Orangutan Optimization Algorithm (OOA) to fine-tune model parameters, enhancing detection accuracy under variable lighting and weather conditions.
- Developed an automated detection system capable of identifying critical Components of the Personal Fall Arrest System (PFAS), including lifelines, harnesses, and helmets.
- Validated the proposed system on both real-world construction sites and controlled laboratory environments to ensure its reliability and generalizability.
- Demonstrated superior results in F1-score, recall, and precision when compared to conventional deep learning models.

Unlike standard YOLOv8 models with generic attention modules, the proposed 2D-3CKO-MTHSAN integrates three complementary strategies: (i) 3-channel Kendall- τ ranking for prioritizing the most informative features, (ii) multi-transform hybrid spatial attention (MTHSAN) to handle multiscale and occluded targets, and (iii) entropy-based preprocessing for enhancing small-scale aerial PPE/PFAS detection. These innovations collectively enable UAV-edge-friendly inference with robustness to motion blur, occlusion, and scale variance — capabilities beyond a simple YOLOv8 + attention baseline.

The remainder of this Section "Literature survey" of the text consists of literature reviews, followed by Section "Proposed Methodology" with suggested approaches, Section "Results and Discussion" with results and discussion, and Section "Conclusions" with a conclusion and future work.

Literature survey

Recent research on drone-assisted safety monitoring has explored a range of neural-network-based methods, yet many studies remain limited in scope or domain. Li et al.¹⁶ developed a 3D Concrete Printing (3DCP) model integrating robotic software and hardware for automated geometry generation, improving construction precision but at a high cost and with limited adaptability to complex sites. Kuwahara et al.¹⁷ proposed a Cluster-based Flight Path Construction Method (CFPCM) using RGB-D sensing and YOLO-based detection; although

effective in agriculture, it must be adapted for dynamic construction scenes. Papyan et al.¹⁸ introduced a CNN-DOA framework for acoustic person detection, useful for distress-signal recognition but sensitive to ambient noise. Cao et al.¹⁹ designed an Efficient Layer Aggregation Network (ELAN) that improves small-object detection (75% IoU) but degrades under occlusion or poor lighting conditions. Li et al.²⁰ proposed a Multi-Task Share-Split-Collaborate Architecture (MTM-SSCA) that enables semi-autonomous drones to achieve high classification accuracy at the cost of higher computational demand. Zhang et al.²¹ used Deep Reinforcement Learning (DRL) for cost-efficient UAV optimization, though training stability and convergence remain challenges. Magaia et al.²² presented an App-based Long-Distance Patient Monitoring system that demonstrates the benefits of UAV-based remote sensing, yet it targets healthcare rather than construction safety.

While these efforts showcase deep learning for UAV perception, most were evaluated in controlled or non-construction domains. They rarely address real-time detection under occlusion, variable altitude, and motion blur typical of aerial safety monitoring. Recent computer vision work^{23–26} on active learning and anomaly detection^{23–26} achieves high accuracy in healthcare or surveillance videos but is not optimized for construction environments characterized by cluttered backgrounds, moving cranes, and small-scale workers.

To further strengthen the case, additional studies highlight key deficiencies in standard detectors such as YOLOv8 and Faster R-CNN. Zhou et al.²⁷ demonstrated that parallel multiscale feature fusion improves structure recovery in overhead scenes, indicating that single-path backbones are inadequate for small, scale-varying aerial targets. Wu et al.²⁸ showed that powerful detectors increase flight power and reduce airtime, underscoring the need for edge-friendly, compact models. Cao et al.²⁹ showed that the risk of falling remains high in mega-project environments, underscoring the need for precise PFAS component detection rather than generic person detection. Li et al.³⁰ found that volatile air-to-ground links limit offloading, requiring lightweight onboard inference. Abbas et al.³¹ found that 2D RGB detectors miss occluded or overlapping subjects, suggesting that hybrid spatial attention or depth cues are essential. Complementary works^{23–26} further support multi-scale fusion, compact model design, and hybrid transformer integration for improved UAV vision. Table 1 summarizes representative methods and their strengths and weaknesses.

Research gap and motivation

Current PPE-monitoring systems based on YOLOv8, Faster R-CNN, or SSD deliver strong results in ground-level imagery but deteriorate under drone perspectives where workers appear small, occluded, and scale-variant. Parallel multiscale fusion²⁷ and edge-energy constraints²⁸ reveal that existing detectors are too heavy for onboard processing and lack temporal adaptability. Moreover, volatile air-to-ground links³⁰ and occlusion effects³¹ limit their deployment in real sites.

Hence, this study proposes a lightweight hybrid-attention detector optimized for UAVs, combining entropy-based preprocessing, τ -Kendall feature selection, and multi-transform fusion to achieve robust real-time PPE and PFAS detection under dynamic construction conditions.

Problem statement

Construction sites involve rapidly changing environments, moving equipment, and multiple workers performing high-risk tasks at different elevations. Existing deep detectors often fail to maintain accuracy under occlusion, motion blur, and illumination variation. There is a pressing need for an adaptive, energy-efficient UAV-based system that can detect PPE and PFAS components in real time to prevent fall-related incidents. This research addresses that gap through a hybrid neural architecture and optimized deployment strategy for drone-assisted safety monitoring.

Proposed methodology

This section presents a new Discrete Dilated Cosine Causal Convolution Krawtchouk Orangutan Multi-Tchebichef Head Self-Attention Network (2D-3CKO-MTHSAN) model to enhance the correctness and trustworthiness of PPE detection. Figure 1 illustrates the block diagram of 2D-3CKO-MTHSAN. The technique uses real-time drone-based observation to continuously acquire visual data at actual building sites and in laboratory simulations, focusing on identifying the main components of the Personal Fall Arrest System (PFAS), including helmets, harnesses, and lifelines. Image enhancement and feature relevance are facilitated by a two-stage preprocessing pipeline comprising entropy-based filtering and τ -Kendall correlation analysis. Adaptive Causal Decision Transformers learn deep spatio-temporal features, and the model parameters DCCMSA-

Reference	Method	Advantages	Limitations
Li et al. ¹⁶	3DCP	Enables automated complex geometry; increases building accuracy	High cost; limited adaptability to rough terrain
Kuwahara et al. ¹⁷	CFPCM	Optimized drone route; RGB-D + YOLO detection	Evaluated only in agriculture; not tuned for dynamic sites
Papyan et al. ¹⁸	CNNs-DOA	Detects human distress acoustically	Limited range; noise sensitivity
Cao et al. ¹⁹	ELAN	Improved small-object detection	Struggles in low-light/occluded scenes
Li et al. ²⁰	MTM-SSCA	High multitask accuracy	Computationally heavy; large training data
Zhang et al. ²¹	DRL	Adaptive UAV optimization	Slow convergence; stability issues
Magaia et al. ²²	App-ALDPMC	Enables long-distance monitoring	Dependent on connectivity and mobile hardware

Table 1. A summary of the approach being assessed.

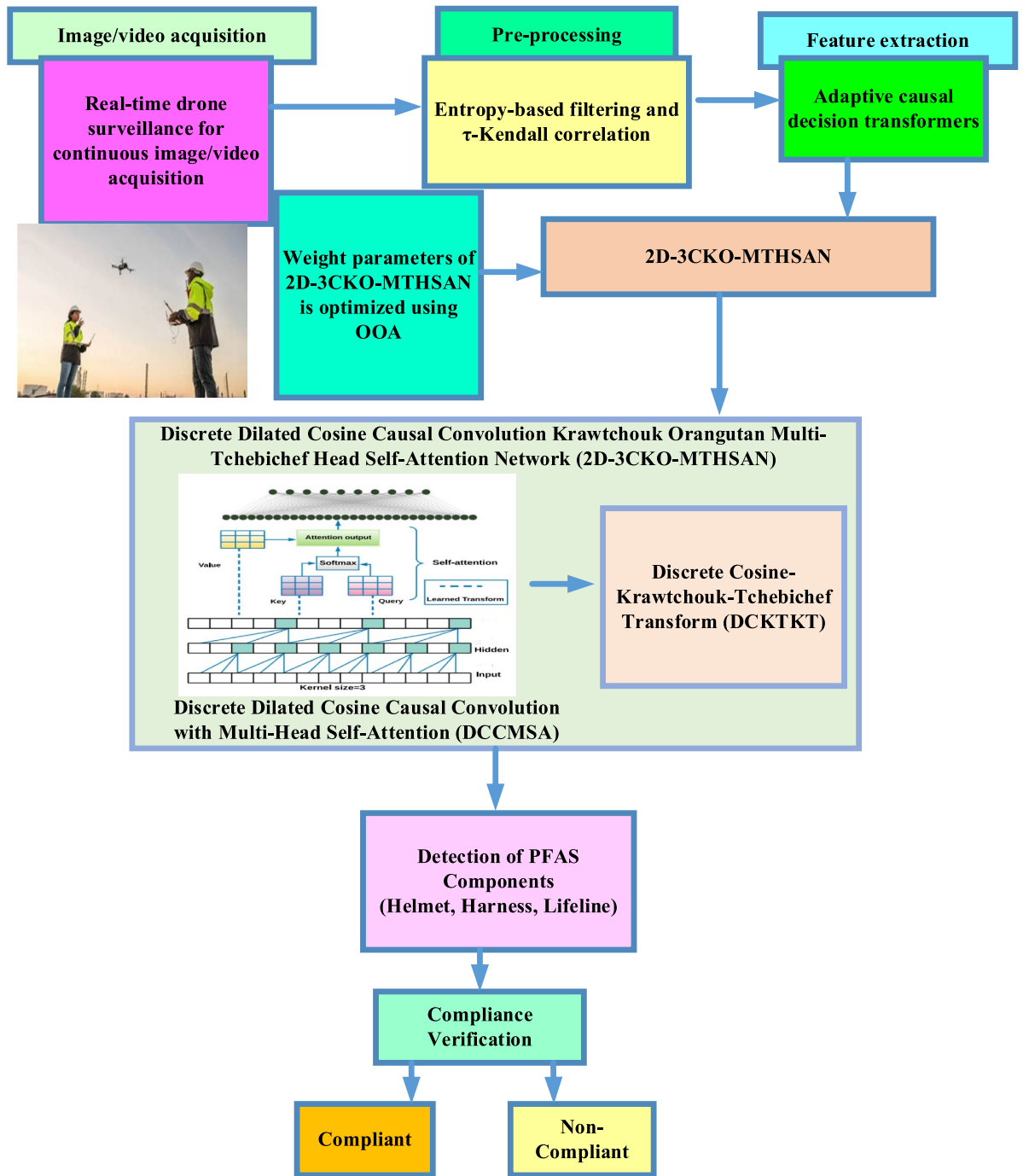


Fig. 1. Block diagram of the 2D-3CKO-MTHSAN framework showing the end-to-end pipeline for real-time drone-based PPE detection and PFAS compliance monitoring.

DCKTKT are optimized by the Orngutan Optimization Algorithm (OOA) to maintain consistent performance under changing environmental conditions.

Figure 1 illustrates the end-to-end process of the proposed 2D-3CKO-MTHSAN framework for real-time drone-based safety monitoring. It begins with real-time drone monitoring to capture images/videos around the clock on construction sites. Images are preprocessed via entropy-based filtering and τ -Kendall correlation analysis to improve their quality and feature relevance. Deep spatio-temporal features are obtained using Adaptive Causal Decision Transformers coupled with Discrete Dilated Cosine Causal Convolution and Multi-Head Self-Attention (DCCMSA). The model parameters are adjusted using the Orngutan Optimization Algorithm (OOA), and the Discrete Cosine-Krawtchouk-Tchebichef Transform (DCKTKT) is used for robust feature representation. Finally, the system detects Personal Fall Arrest System (PFAS) components—helmet, harness, and lifeline—and verifies compliance, identifying workers as compliant or non-compliant.

Data acquisition

A real-time drone monitoring approach is applied in this research to automatically record images and videos of workers performing at-height activities in dynamic construction areas, with the goal of optimizing safety monitoring and reducing accidents caused by the absence of personal protective equipment (PPE). The research employs the publicly available Construction Site Safety Image Dataset²³ on Kaggle, hosted by Roboflow, which contains 2801 YOLOv8-labeled images. The data is split into training (2605 images), validation (114 images), and test (82 images) sets, with corresponding image and label directories. 10 labeled object classes exist: "Mask," "Hardhat," "NO-Hardhat," "NO-Mask," "NO-Safety Vest," "Person," "Safety Cone," "Safety Vest," "machinery," and "vehicle," which allow multi-class PPE and object detection. Each annotated sample in the dataset is pre-equipped with bounding box coordinates and class labels, enabling computer vision models for autonomous PPE compliance monitoring and worker safety observation in construction zones to be developed and tested more conveniently. Next, these input data are fed into the entropy-based filtering and τ -Kendall correlation analysis for enhancing the relevance and coherence of the visual information collected by drones, and this is explained below:

Pre-processing using entropy-based filtering and τ -Kendall correlation analysis

To enhance the informativeness, contextuality, and interpretability of drone-collected visual data, a two-stage preprocessing technique is employed that combines entropy filtering and τ -Kendall correlation analysis²⁴. The procedure eliminates the redundant features, retaining only the most informative and significant ones, thus improving the precision and reliability of PPE detection across various real-world construction site conditions.

Entropy-based filtering

Entropy is a measure of the uncertainty in the distribution of information in a data set. Shannon's entropy hypothesis is used in drone construction monitoring to assess features and retain the most informative features from image data. The process includes the following computational steps:

Step 1: Normalize the Decision Matrix

The decision matrix x_{lk} , where l is the parameter index and k is the attribute index, is normalized using Eq. (1):

$$g_{lk} = \frac{x_{lk}}{\sum_{q=1}^p x_{lk}}, (p = 1, \dots, q, l = 1, \dots, p) \quad (1)$$

where, x_{lk} : Raw pixel intensity or feature value for the l^{th} parameter and k^{th} attribute, g_{lk} : Normalized value, q : Total number of attributes, p : Number of parameters (features).

Step 2: Compute Shannon Entropy for Each Feature

Entropy H_q for feature i is calculated in Eq. (2):

$$H_q = -l \sum_{q=1}^i g_{lk} \text{Bc}g_{lk} \quad (2)$$

where: $l = \frac{1}{\ln m}$: Entropy normalization constant, H_q : Entropy value representing the amount of disorder or uncertainty for the k^{th} feature.

Step 3: Determine the Degree of Uncertainty

The uncertainty or deviation for each feature i is then calculated in Eq. (3):

$$a_q = 1 - H_q \quad (3)$$

where a_q conveys the degree of uncertainty or variation of the data for a characteristic q .

Step 4: Compute Feature Weights

To prioritize more informative features, a normalized weight vector \widehat{R}_q for each attribute is computed using Eq. (4):

$$\widehat{R}_q = \frac{a_q}{\sum_{q=1}^q a_q} \quad (4)$$

where \widehat{R}_q represents the Weight assigned to the j^{th} attribute indicating its relative importance for further processing.

This entropy-based approach ensures that only the features contributing the most meaningful information are retained for downstream analysis.

τ -Kendall correlation analysis

While traditional correlation metrics like Pearson's correlation assume normally distributed data and are sensitive to outliers, construction site imagery often exhibits non-Gaussian noise due to varying lighting, weather, and structural occlusions. The τ -Kendall correlation, therefore, a complementary, rank-based measure of dependency, is used to robustly assess dependencies between visual features.

Given a sample of n observation pairs $(r_1, s_1), (r_2, s_2), \dots, (r_m, s_m)$, Kendall's tau $\widehat{\tau}$ is computed as Eq. (5):

$$\widehat{\omega} = \binom{r}{2}^{-1} \sum_{1 < l < n < m} \text{sgn}[(r_m - r_n)(s_m - s_n)] \quad (5)$$

where the sign function $\text{sgn}(\alpha)$ is defined in Eq. (6):

$$\text{sgn}(\alpha) = \begin{cases} 1 & \text{if } \alpha > 0 \\ 0 & \text{if } \alpha = 0 \\ -1 & \text{if } \alpha < 0 \end{cases} \quad (6)$$

where, m : Number of observations, $\widehat{\omega}$: Kendall's tau correlation coefficient in the range $[-1, +1]$, indicating negative, neutral, or positive correlation, r_m, r_n, s_m, s_n : Ranked feature values for observation m , sgn : Signum function to determine concordant or discordant pairs.

where $n, m = 1, 2, \dots, n$ and $\text{sgn}[\cdot]$ represent a sign function.

This correlation analysis identifies and retains highly interdependent feature pairs, while discarding irrelevant or noisy associations, making it particularly effective in cluttered or occluded construction site images.

Then these pre-processed data are given to the Adaptive causal decision transformers for extracting important features, and their explanations are given below:

Feature extraction using adaptive causal decision transformers for real-time drone-assisted safety monitoring in construction

Following the preprocessing phase, which applies entropy-based filtering and τ -Kendall correlation analysis to enhance the clarity and contextual relevance of drone-captured construction site imagery, the refined data is passed to an Adaptive Causal Decision Transformer (AdaCred)²⁵ for advanced feature extraction. This transformer-based mechanism enables learning spatio-temporal representations crucial for detecting safety compliance indicators, such as helmets, harnesses, and lifelines, in dynamic, partially observable environments.

Theoretical foundation for feature extraction using adaptive causal decision transformers

The input visual data is modeled within the framework of a Partially Observable Markov Decision Process (POMDP), defined in Eq. (7):

$$N = (Q, A, P, R, \delta, T) \quad (7)$$

where: Q : Set of observations (e.g., frames or images from drone footage), A : Set of actions (e.g., detection decisions or transitions between attention states), P : State transition probability function, R : Reward function reflecting model accuracy or prediction relevance, δ : Discount factor for temporal weighting, T : Time horizon.

At each time step $t \in \{1, \dots, T\}$, an agent receives an observation $q_s \in Q$, selects an action $b_s \in B$ based on its prior history $h_{1:s} := (q_{1:s}, a_{1:s-1}) \in H_s$, and receives a reward $r_s \sim R_s(h_{1:s}, a_s)$. The next observation q_{s+1} is sampled from the transition distribution $P(q_1)$.

Latent state representation and feature modeling

To effectively model and extract temporal dependencies, the system assumes that each observation \mathbf{f}_s is generated from a latent state vector $\mathbf{f}_s = (f_1, t, \dots, hd, t)^T$, where d is the dimensionality of the latent space. The latent dynamics, observations, and reward functions are defined as follows:

Latent State Transition Eq. (8):

$$f_{j,s} = h_j(d_j^{f \rightarrow f} \Theta g_{s-1}, d_j^b \cdot a_{s-1}, a_j^w \cdot \mathbf{r}_{s-1}, \in_{j,s}^f), \text{ for } j = 1, \dots, d, \quad (8)$$

where: $f_{j,s}$: j^{th} component of the latent state at time s , $h_j(\cdot)$: Transition function for the j^{th} dimension, $d_j^{f \rightarrow f}$: Binary mask indicating causal dependence on the previous latent state, d_j^b : Coefficient for influence from previous action a_{s-1} : Coefficient for influence from previous reward \mathbf{r}_{s-1} : Gaussian noise capturing stochasticity, Θ : Element-wise multiplication.

Observation Eq. (9):

$$q_s = e(d^{h \rightarrow p} \Theta g_s, \in_s^p), \quad (9)$$

where: $e(\cdot)$: Observation generation function, $d^{h \rightarrow p}$: Causal mask from latent state to observation, \in_s^p : Observation noise is given in Eq. (10):

$$r_s = f(d_w^{f \rightarrow w} \Theta g_{s-1}, d^{b \rightarrow w} \cdot a_{s-1}, \in_s^w) \quad (10)$$

where: $f(\cdot)$: Reward function, $d_w^{f \rightarrow w}$: Causal mask from previous latent state to reward, $d^{b \rightarrow w}$: Influence from prior action, \in_s^w : Reward noise.

Transformer-based feature extraction process

The Adaptive Causal Decision Transformer (AdaCred) utilizes a transformer architecture for modeling the temporal sequence, as given in Eq. (11):

$$K_{reg} = -\lambda_{reg} (\|d^{f \rightarrow p}\|_1 + \|d^{f \rightarrow w}\|_1 + \|d^{f \rightarrow f}\|_1 + \|d^{a \rightarrow f}\|_1 + \|\eta\|_1) \quad (11)$$

where λ_{reg} is a regularization coefficient, and $\|\cdot\|_1$ represents the k_1 -norm, which encourages sparsity. Adjusting the objective function to include this regularization term I permits model training and pruning to happen at the same time. Naturally, the selective elimination of non-compact state variables occurs when this regularization term is present, as entries in d^{\rightarrow} shift between 1 and 0, encouraging a condensed, effective depiction of the causal connections. The observation is contained in each token in the sequence d_j^b , previous action a_{s-1} , and received reward r_{s-1} . The model processes these tokens to identify important features across time, which are crucial for determining whether workers are equipped with PPE under dynamic construction site conditions.

Using multi-head self-attention, the transformer extracts dependencies between distant-in-time features and cuts through ambiguity to better represent latent safety-critical features—enabling precise detection of objects like helmets, vests, and safety cones. The features capture spatial appearance and temporal dynamics, thereby enabling the model to effectively distinguish compliant and non-compliant scenarios.

Then, these extracted features are provided to the 2D-3CKO-MTHSAN model to enable robust and accurate detection of safety equipment, and its explanations are given below:

2D-3CKO-MTHSAN model for robust and accurate detection of safety equipment in real-time drone-assisted safety monitoring in construction

He 2D-3CKO-MTHSAN model differs fundamentally from simply augmenting YOLOv8 with attention modules. The 3-channel Kendall- τ operation ranks feature maps by discriminative importance, MTHSAN fuses multiscale and spatially aware features adaptively, and entropy preprocessing emphasizes small or low-contrast PPE/PFAS targets. This combination ensures that UAV-based inference remains lightweight, robust, and accurate in dynamic construction environments—conditions where a plain YOLOv8 with attention modules often fails. This novel framework integrates the strengths of the Discrete Dilated Cosine Causal Convolution with Multi-Head Self-Attention (DCCMSA)²⁶ and Discrete Cosine-Krawtchouk-Tchebichef Transform (DCKTKT)²⁸ to enable robust and accurate detection of safety equipment, and model DCCMSA-DCKTKT parameters are optimized via the Orangutan Optimization Algorithm (OOA)²⁹, and explanations are provided below:

DCCMSA for robust and accurate detection of safety equipment in real-time drone-assisted safety monitoring in construction

Once critical spatio-temporal features have been extracted from preprocessed construction imagery using Adaptive Causal Decision Transformers, these features are passed to the 2D-3CKO-MTHSAN model, which incorporates a DCCMSA mechanism. This stage is designed to enhance the detection of essential safety equipment, such as helmets, harnesses, and lifelines, by efficiently capturing multi-scale contextual dependencies and temporal correlations.

Dilated causal convolution for sequential dependency modeling

In safety-critical video analysis, the temporal sequence of visual features is of paramount importance. Causal convolutions ensure that the model's output at time step t depends only on current and past information, thereby preserving the temporal ordering of inputs. To overcome the limitations of standard convolutions for enlarging the receptive field, dilated convolutions, causal convolutions, and deep networks are used, or huge filters are introduced to expand the receptive field exponentially without significantly increasing computational overhead.

The DCC for a one-dimensional input sequence $x \in R^m$ with filter $f : \{0, m-1\} \rightarrow R$, applied at position t , is defined in Eq. (12):

$$DCC(y * bc)(d) = \sum_{j=0}^{m-1} c(i) \cdot d_{t-c \cdot j} \quad (12)$$

where b is the dilation factor, determining the stride between sampled elements, n is the filter size, and $t - c \cdot j$ shows the index of the input sequence in the past (due to causal nature).

The dilation factor b is defined for each convolutional layer $m = 1, \dots, N$, where N is the total number of convolutional layers. To exponentially increase the receptive field, the dilation factor follows in Eq. (13):

$$a \in [2^0, 2^1, 2^2, \dots, 2^{N-1}] \quad (13)$$

This exponential dilation allows each layer to access a progressively larger temporal context, enabling the detection of long-term dependencies in safety equipment usage patterns without requiring deep network structures.

Multi-head self-attention for temporal importance modeling

Following dilated causal convolution, the Multi-Head Self-Attention (MHSA) mechanism is employed to identify and amplify the significance of critical time steps in the temporal sequence of features. This is especially vital in real-time drone-assisted safety monitoring, where certain frames may carry more contextual information regarding the presence or absence of personal protective equipment (PPE).

The self-attention mechanism operates using three trainable linear projections: queries (P), keys (Q), and values (R), where:

The scaled dot-product attention is computed in Eq. (14):

$$f_{sd}^{(f_i)}(P, Q, R) = \text{soft max} \left(\frac{W.M^S}{\sqrt{a_m}} \right) R \quad (14)$$

where: $W.M^S$: Measures similarity between time steps, $\sqrt{a_m}$: Scaling factor to stabilize gradients, *soft max*: Converts similarity scores into attention weights.

Multiple attention heads are computed in parallel to capture diverse temporal patterns. The final multi-head output is given in Eq. (15):

$$Q_{fb} = R_0.\text{concat} \left(f_{sd}^{(f_1)}, \dots, f_{sd}^{(f_{m-1})}, f_{sd}^{(f_m)} \right) \quad (15)$$

where: $f_{sd}^{(f_m)}$: Output of the sd^{th} attention head, R_0 : Learned projection matrix to map concatenated heads back to input dimension.

The result is a refined sequence of attention-weighted features that emphasize critical frames where safety equipment is more visibly distinguishable or functionally relevant.

Integration for efficient and accurate detection

The combination of multi-head self-attention and dilated causal convolution enables the 2D-3CKO-MTHSAN model to efficiently process feature-rich drone footage with both local and global temporal awareness. This hybrid architecture:

- Avoids reliance on recurrent neural networks, enabling faster training and inference
- Preserves input resolution and sequence integrity
- Focus computational resources on the most informative temporal segments

These features collectively lead to superior recognition accuracy for detecting PPE components in construction environments.

By integrating Discrete Dilated Cosine Causal Convolutions and Multi-Head Self-Attention (DCCMSA), the system is efficient in detecting and situating PPE visibility in temporally changing drone videos. This directly enables safe, scalable, and real-time safety inspections on dynamic building sites, with enhanced detection accuracy, reduced computational burden, and quicker response times in hazardous work environments.

Then, in order to improve performance, the DCCMSA was integrated with DCKTKT in order to forecast the longevity and competitiveness of small firms, supporting strategically driven entrepreneurship, and the reasons are given below:

DCKTKT for accurate detection of safety equipment in real-time drone-assisted safety monitoring

In live construction-site drone-based safety monitoring, effective feature extraction from aerial photographs is a key factor for correctly identifying personal protective equipment (PPE) used by site workers. To address this requirement, a hybrid transform, DCKTKT (Discrete Cosine-Krawtchouk-Tchebichef Transform), is applied to extract distinguishing features from drone-mounted camera frames.

DCKTKT transform takes the benefits of three orthogonal polynomial-based transforms:

- **Discrete Cosine Transform (DCT)** to compact energy,
- **Krawtchouk Polynomials (KP)** to localize the spatial domain efficiently, and
- **Tchebichef Polynomials (TP)** to improve texture and shape representation.

This composite transform facilitates robust analysis of challenging construction site images by emphasizing key spatial textures while excluding redundant background clutter.

Let P denote the size of the input image block, and let $p = 0, 1, \dots, F - 1$. The hybrid orthogonal polynomial transform of order q , denoted as $W_t(q, F)$; $M_i(F)$, is defined in Eq. (16):

$$W_t(q, F) = \sum_{i=0}^{F-1} M_i(F) N_i(q; F) W_i(f; F) N_i(q; F) W_i(f; F), f, q = 0, 1, \dots, F - 1 \quad (16)$$

where $N_i(q; F)$, $N_i(t; T)$, and $M_i(F)$ are krawtchouk polynomial (KP), tchebichef polynomial (TP) and discrete Cosine transform (DCT), respectively.

This combination results in an enhanced transformation function that efficiently captures both global and local structural variations in construction site imagery.

Matrix formulation

The DCKTKT transform in matrix form is represented as Eq. (17–18):

$$A = M_A M_S M_D M_S M_D \quad (17)$$

$$A = M_A (M_S M_D)^2 = A_{\text{DCKTKT}} \quad (18)$$

where M_S , M_A and M_D are matrix form of KP, DCT, and TP, respectively.

This construction allows hierarchical encoding of image content, with the energy compacted in the initial coefficients. The **top-left quadrant** of the DCKTKT-transformed matrix contains most of the signal energy, indicating efficient information representation and compression.

A 2D DCT is applied across image patches or the full frame to extract spatial–frequency-domain features. To see this, consider a 128×128 image with parameter. The informative energy of most of this image is concentrated in the top-left corner of the resulting image. These coefficients represent compressed descriptors that preserve important structural and contextual information about the occurrence and location of safety equipment such as helmets, vests, and gloves.

Integration of DCCMSA-DCKTKT model

These features from the DCKTKT transformation are then fed into the DCCMSA-DCKTKT model. This deep learning model is fine-tuned to process the spatial representations that are compressed and maintain both:

- **Channel-wise correlations** across multiple orthogonal feature maps,
- **Temporal-spatial dependencies** from drone video sequences.

The hybrid attention mechanism enables dynamic focus on relevant regions (e.g., workers wearing PPE) while suppressing background clutter and irrelevant motion. This pipeline enables precise, real-time detection of safety equipment under various conditions, including illumination, occlusions, and worker motion.

Benefits of DCCMSA-DCKTKT in drone-aided environments

- **High Energy Compaction:** Effective utilization of transmission bandwidth and lowered computation load on edge devices.
- **Strong Localization:** Increases the separability of features pertaining to PPE.
- **Noise Suppression:** Reduces the effect of ambient noise and motion blur typical of drone imagery.

The weight parameters are then used to reduce computational time, computational complexity, and error rate while increasing accuracy η of DCCMSA-DCKTKT is optimized using Orangutan Optimization Algorithm (OOA) for Accurate Detection of Safety Equipment in Real-Time Drone-Assisted Safety Monitoring and its explanations are given below:

Orangutan Optimization Algorithm (OOA) for optimizing DCCMSA-DCKTKT parameters for accurate detection of safety equipment in real-time drone-assisted safety monitoring

To achieve high accuracy, low computational complexity, and robust generalization in real-time construction site monitoring with drone-supported imagery for safety equipment detection, the DCCMSA-DCKTKT's weight parameters are optimized using OOA. This bio-inspired metaheuristic algorithm models the dynamic foraging process and adaptability of orangutans in their natural habitat to enable productive exploration and exploitation in high-dimensional optimization spaces.

Overview of the OOA-based optimization framework

The DCCMSA-DCKTKT model incorporates spatio-temporal feature extraction, cross-modal attention, and adaptive transformer encoders for processing drone-acquired visual streams. Parameter tuning of key components such as attention heads, convolutional weights, and kernel transformation matrices is essential to minimize false detections and improve real-time performance. The OOA is used to optimize parameters based on simulated natural foraging patterns of orangutans, which naturally explore varied areas in search of optimal food.

Figure 2 shows the step-by-step process of OOA for Optimizing DCCMSA-DCKTKT Parameters for Accurate Detection of Safety Equipment in Real-Time Drone-Assisted Safety Monitoring, and its steps are mentioned as follows:

Step 1: Initialization of Orangutan Agents in Parameter Space

At the first stage of the Orangutan Optimization Algorithm (OOA), a population of orangutan agents is initialized randomly in a high-dimensional parameter space, and every agent is a possible solution vector with the tunable parameters of the DCCMSA-DCKTKT model encoded. The parameters are convolutional kernel weights, self-attention coefficients, multiscale fusion weights, and temporal transfer thresholds. Each orangutan agent is assigned to a specific position within the parameter space, providing ample coverage and diversity throughout the search space. This enables the algorithm to search multiple regions of the solution space concurrently, increasing the likelihood of finding a globally optimal configuration for robust, real-time drone-assisted safety surveillance in construction settings. Initialization ensures that all parameters are within specified limits, allowing the algorithm to begin optimization without violating constraints or encountering convergence issues.

Step 2: Random Generation of Candidate Solutions

In this step, candidate solutions are generated randomly from the positions of the initialized orangutan agents to adequately explore the search space. Every candidate is a slightly different version of a current agent, adding diversity to the population. This is done by applying stochastic perturbations to the current parameter values of each orangutan —i.e., the convolutional weights, attention scores, and transfer coefficients in the DCCMSA-DCKTKT model. The goal is to promote exploration by mimicking uncertain foraging activity, enabling the algorithm to draw samples from various configurations that might yield better detection accuracy.

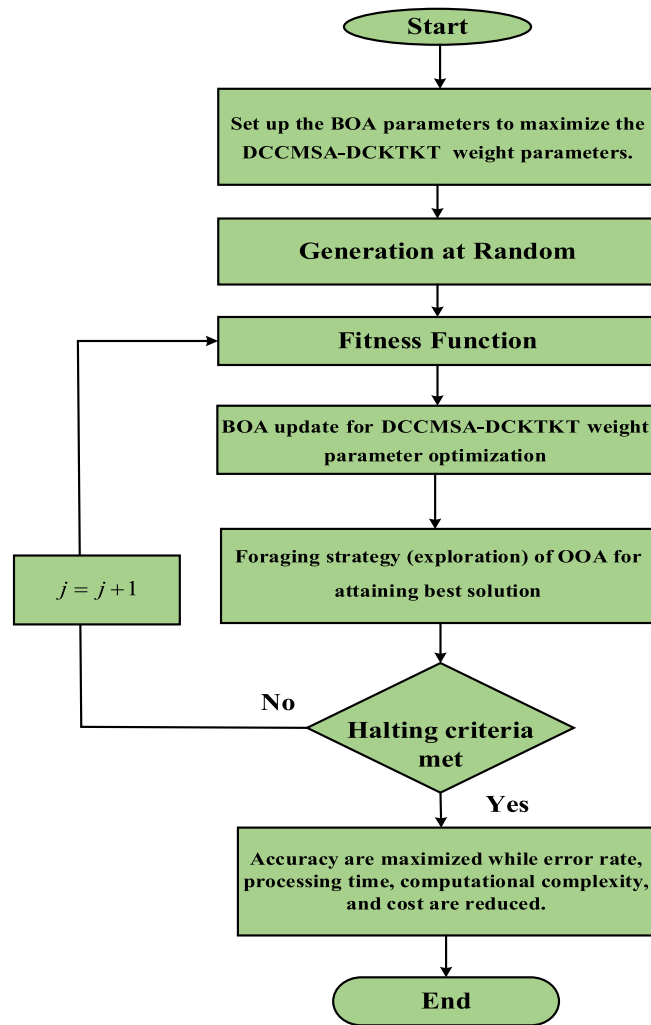


Fig. 2. Stepwise flow of the Orangutan Optimization Algorithm (OOA) for tuning DCCMSA-DCKTKT parameters to maximize safety equipment detection accuracy.

These randomly generated candidates are evaluated for fitness based on their ability to enhance real-time drone-assisted detection of safety equipment, and only those showing potential for improvement are retained for further optimization.

Step 3: Fitness Function Evaluation

A multi-objective fitness function *Fitness function* is evaluated for each agent to guide the optimization, considering prediction accuracy, computational time, and resource efficiency. The function is formulated in Eq. (19):

$$Fitness\ function = Optimize(\eta) \quad (19)$$

Step 4: Foraging strategy (exploration) of OOA for attaining best solution

Orangutans are known for their extensive foraging behavior, which takes them across vast areas to find food. In the OOA framework, this translates into a wide exploration of the search space. The set of superior solutions GC_j orangutan is defined in Eq. (20):

$$GC_j = \{N_m : G_m < G_j \text{ and } n \neq j\} \quad (20)$$

Here GC_j is the set of candidate food sources' locations for the j th orangutan, N_m is the is the orangutan with a better objective function value than j th orangutan, and G_m is the its objective function value.

This equation balances global exploration and local refinement, allowing agents to traverse towards better-performing configurations while maintaining solution diversity.

Step 5: Iterative Update and Optimization

The iteration continues until a convergence criterion is met—either:

- Maximum iterations reached, or

- Minimal change in fitness value over successive generations

Advantages of OOA-Optimized DCCMSA-DCKTKT in drone-assisted PPE detection

- **Accuracy Enhancement:** Parameter refinement leads to higher true positive rates in detecting critical safety equipment under diverse lighting and occlusion scenarios.
- **Computational Efficiency:** Redundant parameters are eliminated, reducing GPU usage and processing latency.
- **Scalability and Adaptability:** The model generalizes effectively across different construction site layouts and environmental conditions.
- **Stability Under Dynamic Conditions:** Maintains detection consistency in motion-blurred, noisy, or crowded drone feeds.

By using the Orangutan Optimization Algorithm to tune the DCCMSA-DCKTKT model, the real-time drone-assisted construction site safety monitoring system achieves high-precision PPE detection. The biologically inspired search and convergence mechanisms that ensure the model is lightweight, precise, and robust help meet the operational requirements of active-safety compliance systems implemented in construction areas. For reproducibility and clarity, the hyperparameter settings and implementation details of the DCKTKT-OOA framework presented below are outlined. The DCKTKT module uses five convolution layers (kernel size = 3×3 , stride = 1, dilation = 2), followed by two fully connected layers containing 256 and 64 neurons, respectively, and a SoftMax classifier for PPE category prediction. The Orangutan Optimization Algorithm (OOA) is configured with a population size of 30 agents, a maximum of 100 iterations, an exploration–exploitation balance factor of 0.6, and a fitness tolerance of 1×10^{-5} . The learning rates are set to 1×10^{-3} for the backbone network, 5×10^{-4} for the DCKTKT fusion layers, and 1×10^{-5} for the OOA fine-tuning stage. Convergence is typically achieved after ~ 84 iterations, ensuring stable training without overfitting. These parameters have been optimized through extensive trials to balance computational efficiency and detection accuracy in real-time drone-assisted PPE monitoring. Then the results and discussions are given below:

Results and discussion

This section presents the results and discussion of 2D-3CKO-MTHSAN for Real-Time Drone-Assisted Safety Monitoring in Construction.

Dataset descriptions

In this study, a real-time drone surveillance approach is employed to continuously acquire images and videos of workers performing at-height tasks in dynamic construction environments, aimed at enhancing safety monitoring and preventing accidents caused by the absence of personal protective equipment (PPE). The analysis utilizes the publicly available Construction Site Safety Image Dataset from Roboflow, hosted on Kaggle, consisting of 2801 annotated images in YOLOv8 format. The dataset is organized into training (2605 images), validation (114 images), and test (82 images) subsets, each containing the corresponding image and label folders. It includes 10 annotated object classes: 'Hardhat', 'Mask', 'NO-Hardhat', 'NO-Mask', 'NO-Safety Vest', 'Person', 'Safety Cone', 'Safety Vest', 'machinery', and 'vehicle', enabling multi-class PPE and object detection. Each labeled instance within the dataset comprises bounding box coordinates and class identifiers, facilitating the development and evaluation of computer vision models for automated PPE compliance monitoring and worker safety assurance in construction zones. Among them, 80% of the data is used for training and 20% for testing. Table 2 lists the precise parameters used for implementation.

Performance metrics

The suggested 2D-3CKO-MTHSAN method's performance is contrasted with that of the current approaches, including 3DCP¹⁶, CFPCM¹⁷, CNNs-DOA¹⁸, ELAN¹⁹, MTM-SSCA²⁰, DRL²¹, and App-ALDPMC²², respectively, using performance metrics like mistake rate, recall, f1 score, accuracy, precision, Train time, computational complexity, processing time, Hamming loss, Root mean square error (RMSE), mean squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) analysis. Table 3 provides the equations for the performance metrics:

True Positive TS: When the drone detects a genuine safety concern, such as a worker not wearing a hard hat, it may promptly address it and help avoid mishaps.

False Positive DS: When a vest is mistaken for missing equipment, the drone misidentifies the issue and sends out needless alarms.

Parameters	Description
Proposed Neural Network	2D-3CKO-MTHSAN
OS	Windows 10
Optimization	OOA
Dataset	Construction Site Safety Image Dataset
Software	Python 3.7

Table 2. Implementation Parameters.

Performance metrics	Equation (21–25)
Accuracy	$\frac{TS+TF}{TS+TF+DS+DF}$ (21)
Precision	$\frac{TS}{TS+DS}$ (22)
Recall	$\frac{TS}{TS+DF}$ (23)
F1-Score	$2 \times \frac{Precision \times Recall}{Precision + Recall}$ (24)
Specificity	$\frac{TF}{TF+DF}$ (25)

Table 3. Performance metrics.

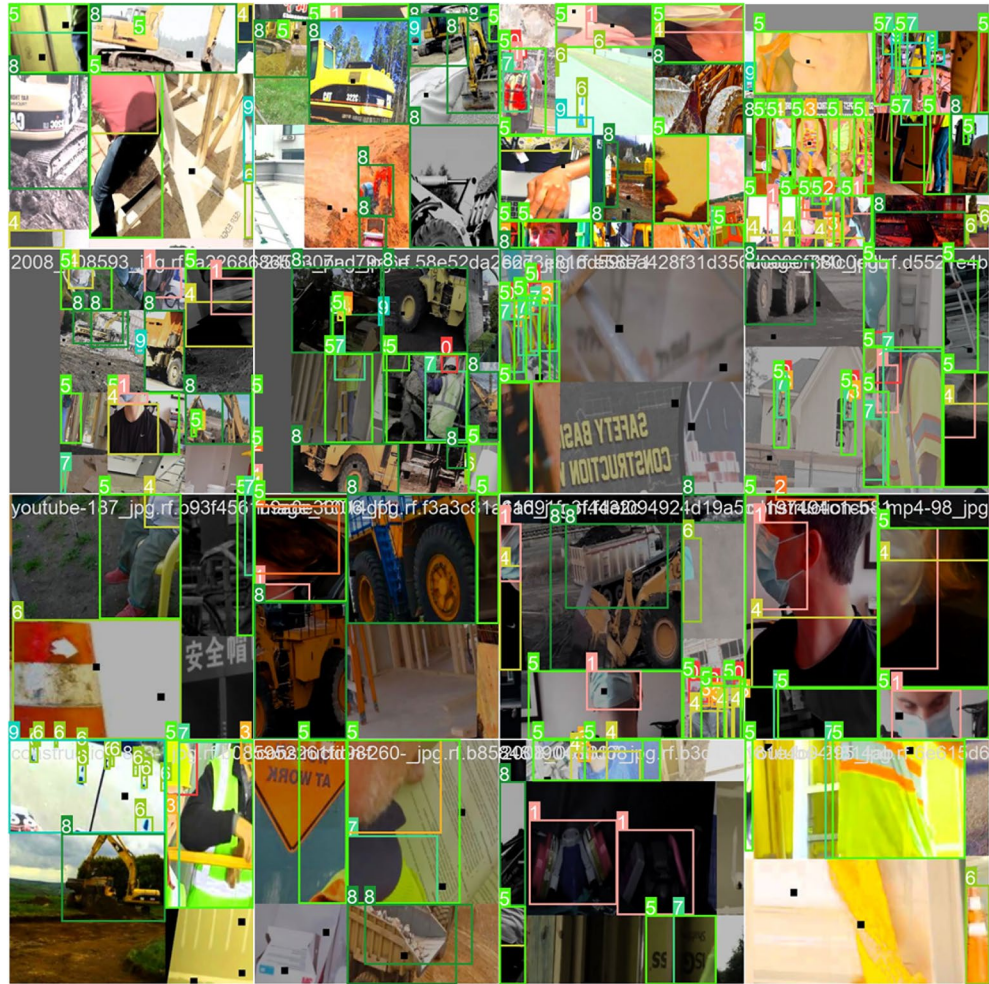


Fig. 3. Representative object detection results on construction site images highlighting helmets, harnesses, and other PPE components.

False Negative DF: Accidents could result from the drone missing a genuine threat, such as someone entering a prohibited area.

True Negative TF: By accurately ignoring safe conditions, the drone prevents false warnings and maintains order. Table 3 shows the Performance metrics below.

Performance analysis

The performance analysis of 2D-3CKO-MTHSAN is discussed here:

Figure 3 shows the Object Detection in Construction Site Images. An object detection model was used to apply bounding boxes and confidence scores to a number of construction-related scenes. The model recognizes several safety components, including employees, machines, vests, and helmets. Different classes or detection

statuses are probably represented by the green and pink boxes. This graphic illustrates how the model can automatically verify safety compliance across a range of construction settings.

Figure 4 shows the Automated Detection of Safety Gear and Equipment in Diverse Work Environments. The object detection outputs photos of industrial and construction sites, with bounding boxes emphasizing machinery, workers, and safety equipment. Color-coding is used to indicate different classifications and levels of confidence. The system demonstrates effective multi-object recognition in challenging environments, supporting hazard detection across a variety of situations and lighting conditions, as well as real-time safety compliance monitoring.

Figure 5 illustrates the Visual Detection of Construction Site Entities. The outputs of this item detection are utilized in numerous building applications. Green, red, and yellow bounding boxes highlight certain features, such as people, equipment, safety equipment, and warning signs, along with their respective confidence scores. The technology proves its mettle by effectively capturing multiple viewpoints and environments to sustain site safety and operational awareness in dynamic, populated work environments.

Figure 6 shows the (a) Distribution of epoch and (b) Distribution of train/box_loss. The training epochs are distributed quite uniformly, with a range of 0–100 in Fig. 6(a). The comparable frequency of data points in each bin suggests that epochs were sampled uniformly throughout the training procedure. The distribution of the train/box_loss data is right-skewed. As the model improves over time, the majority of values cluster around 0.8–0.9 in Fig. 6(b), with fewer examples having higher loss values above 1.1.

Figure 7 shows the (a) Distribution of train/cls_loss and (b) Distribution of train/df_l_loss. Distribution of training-related classification loss. The majority of values fall between 0.5 and 1.0, and as loss increases, frequency decreases in Fig. 7(a). This suggests that there were fewer cases of high classification loss and that the model typically performed well. The distribution of objectness loss. It peaks between 1.1 and 1.2 and then progressively drops toward higher loss values in Fig. 7(b), indicating that while most objectness predictions were reasonably accurate, some had higher losses, suggesting occasional prediction errors.

Figure 8 shows the (a) Distribution of metrics/precision (B) and (b) Distribution of metrics/recall (B). Distribution of accuracy values, where the majority fall within the 0.8–0.9 range. This suggests the model has

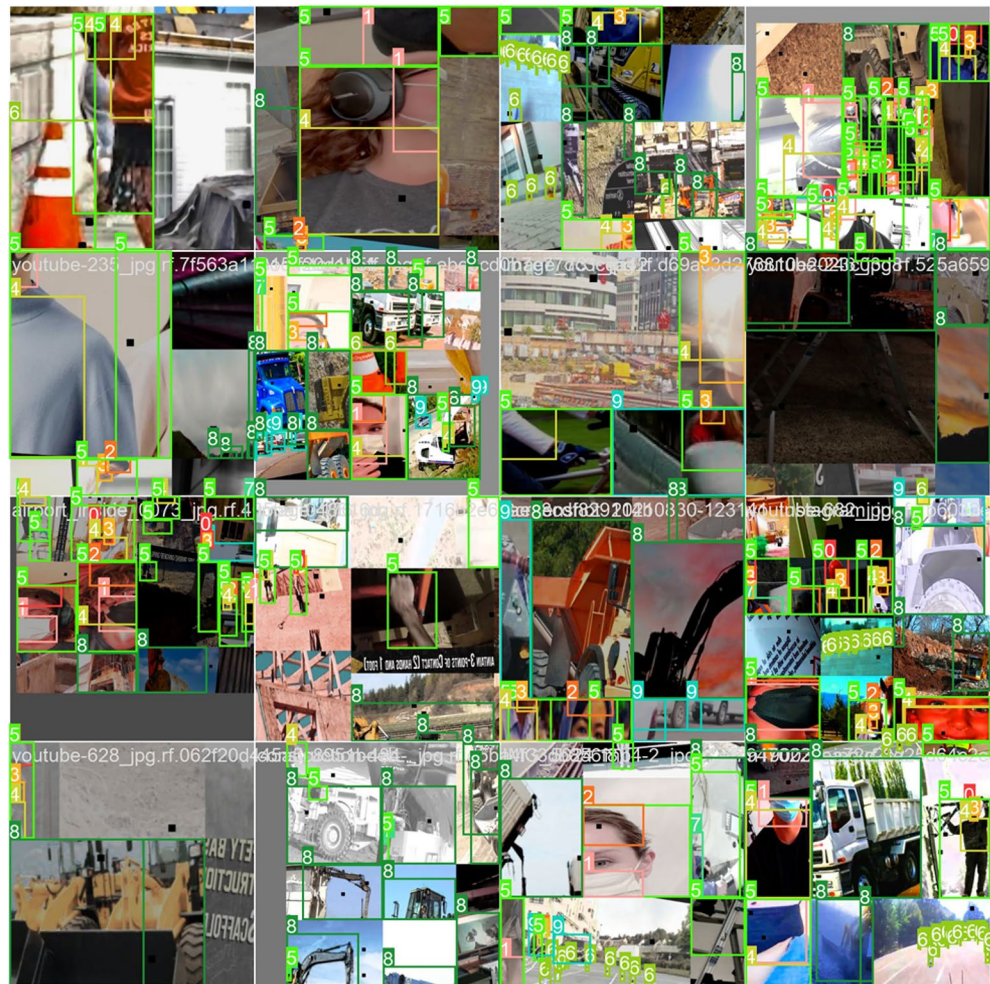


Fig. 4. Automated identification of safety gear across diverse construction site environments, demonstrating robustness to occlusion and variable lighting.

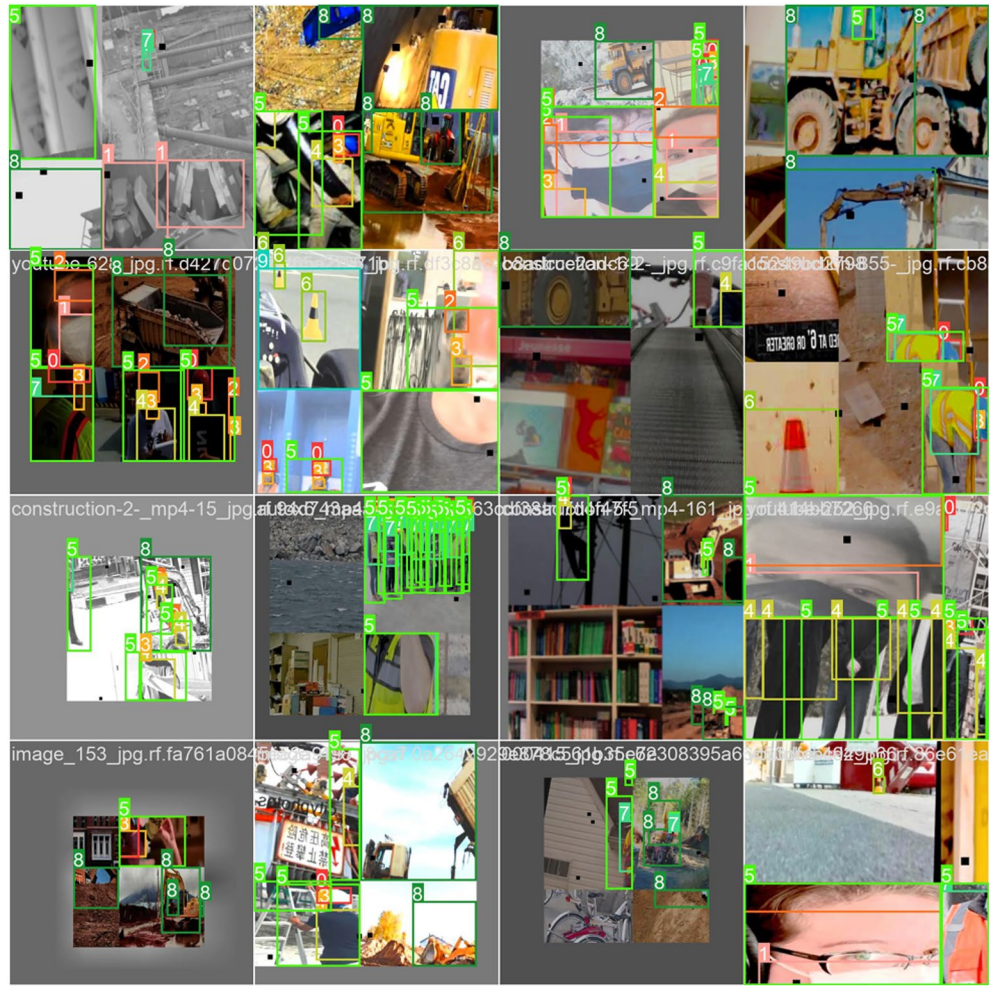


Fig. 5. Visual detection of construction site entities, including workers, vehicles, and equipment, validating multi-class recognition capabilities.

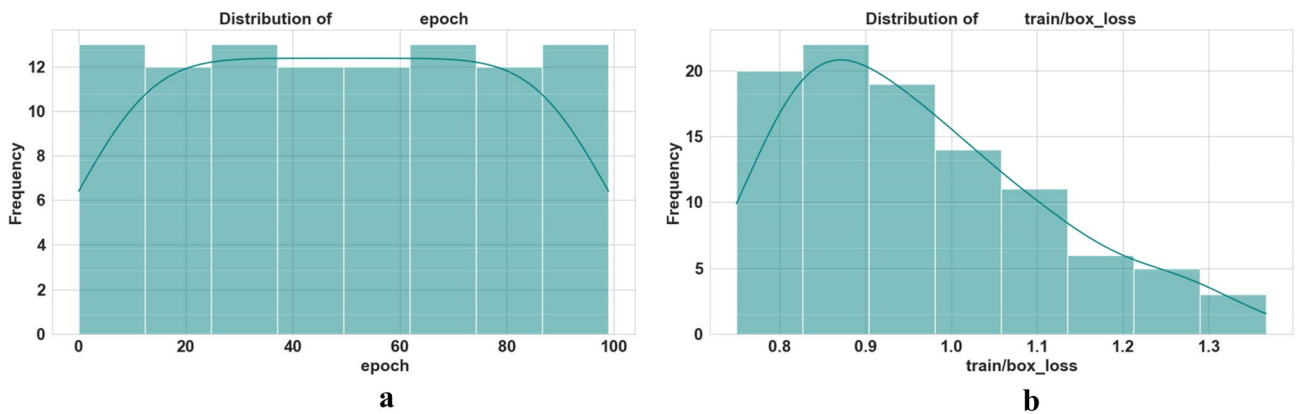


Fig. 6. (a) Epoch distribution, (b) Training box loss trends showing convergence behavior of the 2D-3CKO-MTHSAN model.

high precision, meaning it rarely produces false-positive predictions. Very few numbers are less than 0.6 in Fig. 8 (a). Recall showed a similar trend, with the majority of values falling between 0.6 and 0.7 in Fig. 8 (b). Despite recall's inferiority to accuracy, the upward trend suggests progress over time, indicating improved detection of relevant events (fewer false negatives).

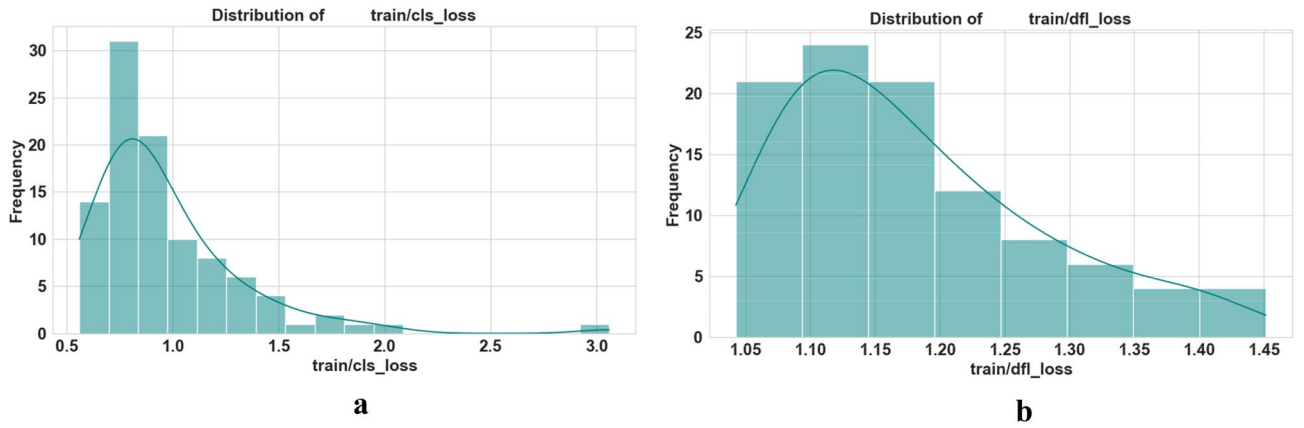


Fig. 7. (a) Training classification loss, (b) Distribution of DFL loss illustrating model’s learning of object localization and class prediction.

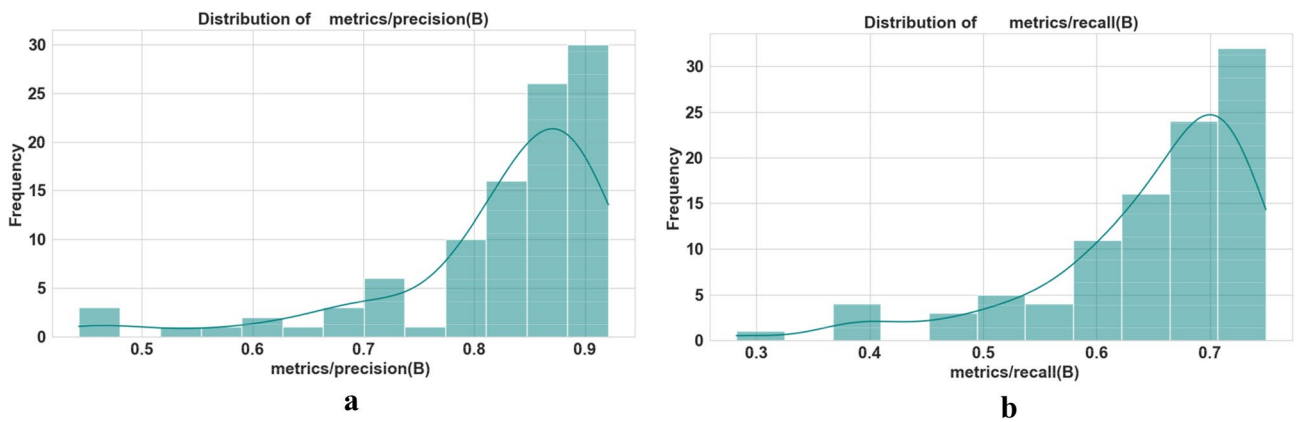


Fig. 8. (a) Precision, (b) Recall trends during training highlighting the trade-offs and overall detection performance across classes.

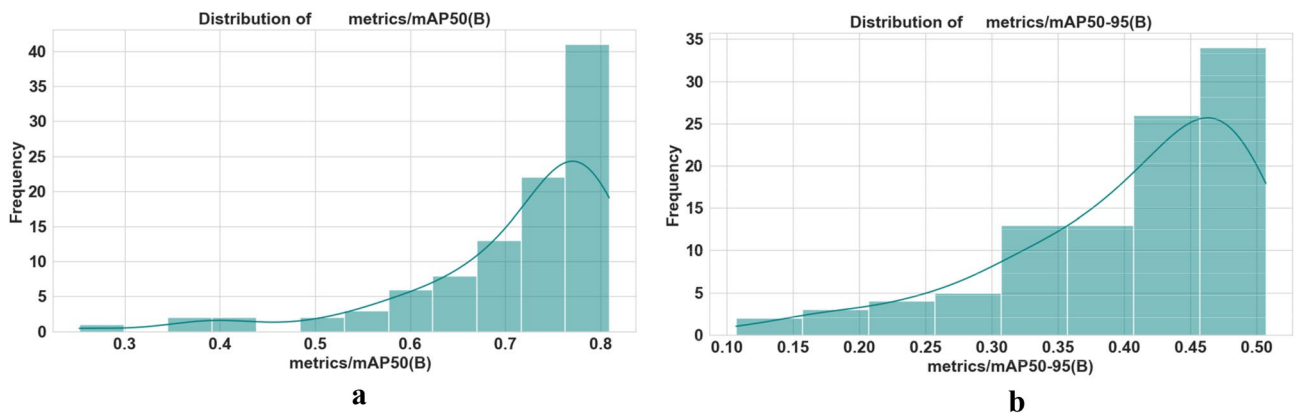


Fig. 9. (a) mAP50, (b) mAP50-95 metrics over training showing detection accuracy at different IoU thresholds for PPE components.

Figure 9 shows the (a) Distribution of metrics/mAP50 (B) and (b) Distribution of metrics/mAP50-95 (B). Given that most data cluster between 0.7 and 0.8, the mAP50 values in Fig. 9(a) are right-skewed, suggesting high detection accuracy at a 50% IoU threshold. The mAP50-95 values in Fig. 9 (b) exhibit modest performance over more stringent IoU thresholds, with a more dispersed peak around 0.45 to 0.50. These figures together

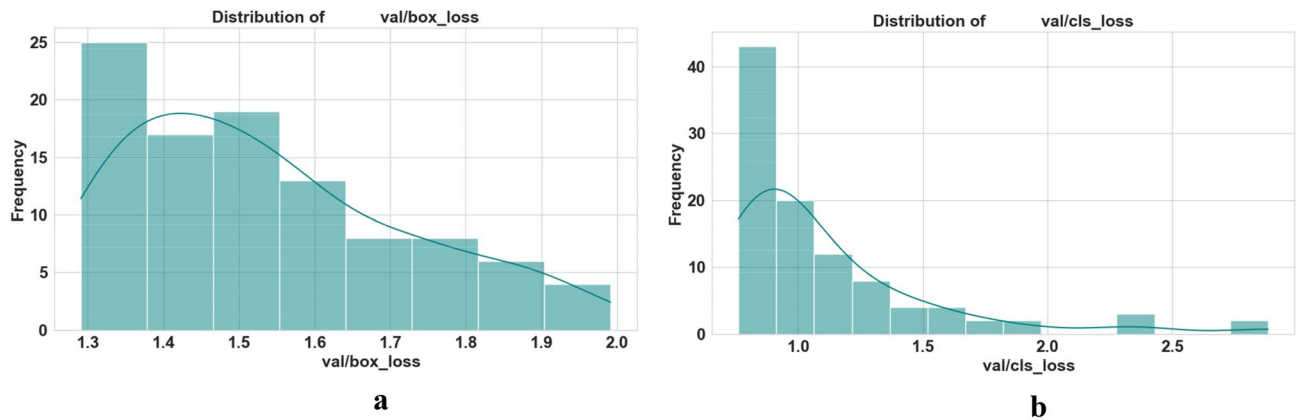


Fig. 10. (a) Validation box loss, (b) Validation classification loss demonstrating model generalization on unseen construction site images.

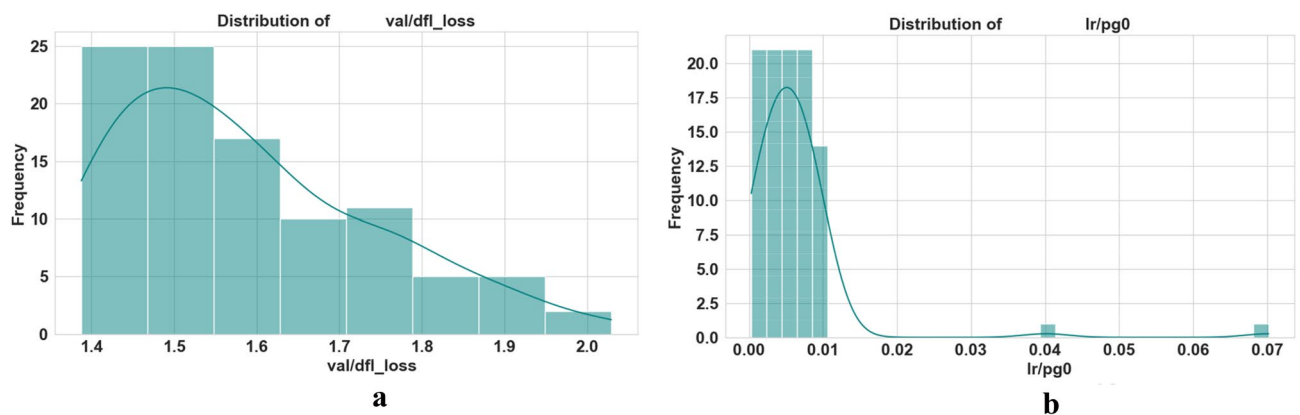


Fig. 11. (a) Validation DFL loss, (b) Layer-wise learning rate progress tracking for stable convergence of 2D-3CKO-MTHSAN.

indicate that the model does well at loose localization but could be strengthened at more stringent detection criteria.

Figure 10 shows the (a) Distribution of `val/box_loss` and (b) Distribution of `val/cls_loss`. The `val/box_loss` values in Fig. 10(a) are primarily between 1.3 and 1.6, indicating significant localization problems with a right-tailed distribution. With most values close to 1.0 and a few outliers above 2.0, Fig. 10 (b) shows a stronger right skew for `val/cls_loss`, suggesting generally acceptable classification with sporadic high-error situations. When taken as a whole, these graphs indicate reliable validation performance, with occasional higher losses that might be improved with additional adjustments.

Figure 11 shows the (a) Distribution of `val/dfl_loss` and (b) Distribution of `lr/pg0`. With most values falling between 1.4 and 1.6, `val/dfl_loss` in Fig. 11 (a) is somewhat skewed to the right, suggesting a reasonable level of object detection performance with a few higher-error occurrences. A conservative learning rate approach is suggested by the distribution in Fig. 11(b), which is heavily skewed toward lower values and peaks near 0.01. When combined, these distributions show a well-managed training procedure with steady validation results and conservative learning rate modifications.

Figure 12 shows the (a) Distribution of `lr/pg1` and (b) Distribution of `lr/pg2`. Demonstrate almost identical distributions between 0.001 and 0.009, suggesting uniformly applied or evenly sampled learning rates during training. To ensure balanced weight updates across the network and possibly promote consistent convergence during training, this uniformity implies a stable, systematic learning rate for all model components.

Figure 13 is the (a) Epoch and (b) `Train/box_loss`. The epoch values in Fig. 13(a), which cluster largely between 25 and 75 for 100 training epochs, with a median of approximately 50, indicate a well-balanced distribution of training rounds. By comparison, the values of `train/box_loss` in Fig. 13(b) range from 0.85 to 1.12, with a median of nearly 1.0. This means that while the outliers may be at most ~ 1.3 , the majority of the box losses remain at a similar level. This suggests that the box regression loss is constant over that small range and that training proceeds similarly across epochs.

Figure 14 shows the (a) F1-Confidence Curve and (b) `Train/cls_loss`. The image displays that training loss behavior and prediction confidence affect model performance. Strong detection performance is indicated by the

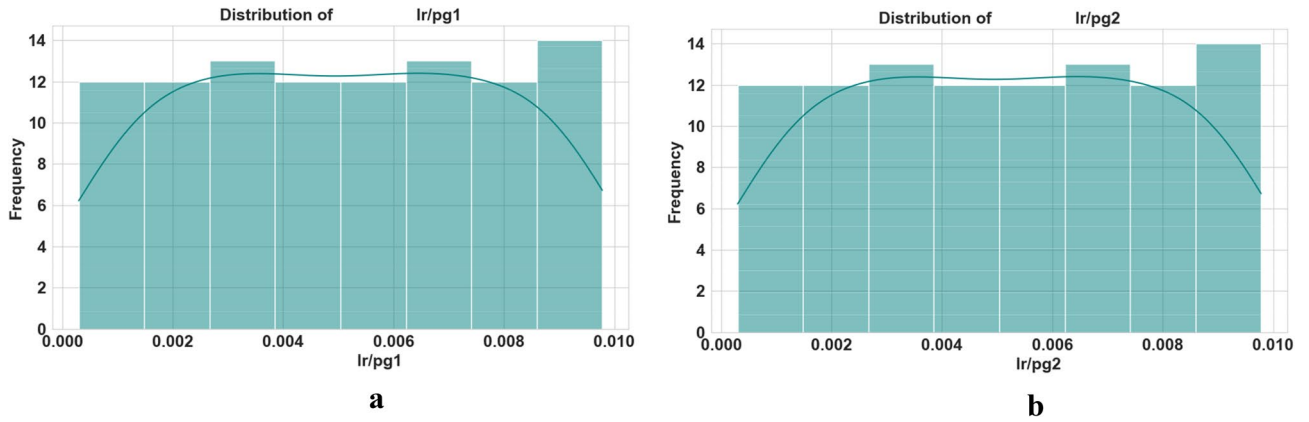


Fig. 12. (a) Learning rate schedule for pg1, (b) Learning rate schedule for pg2 visualizing adaptive optimization over epochs.

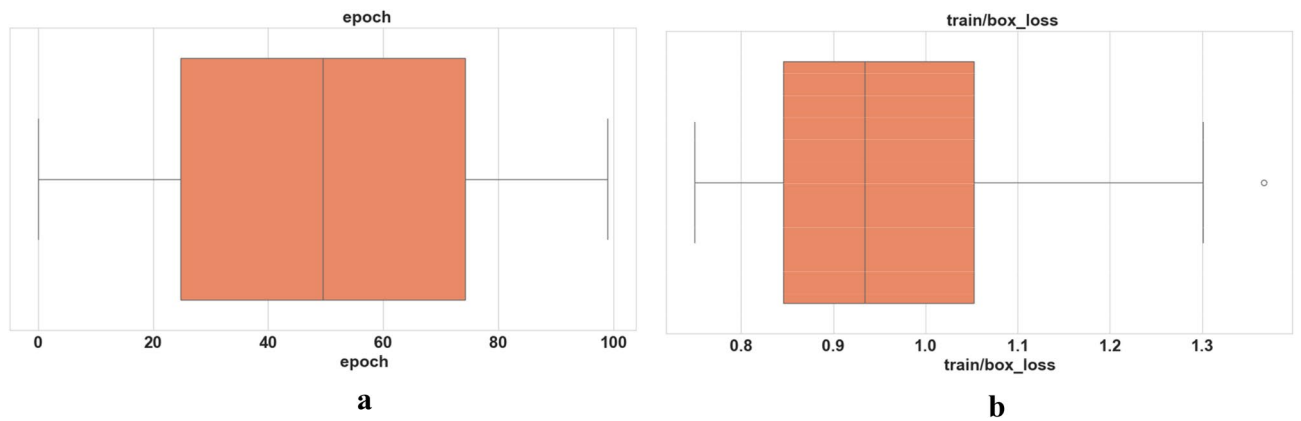


Fig. 13. (a) Epoch progression, (b) Training box loss highlighting convergence dynamics.

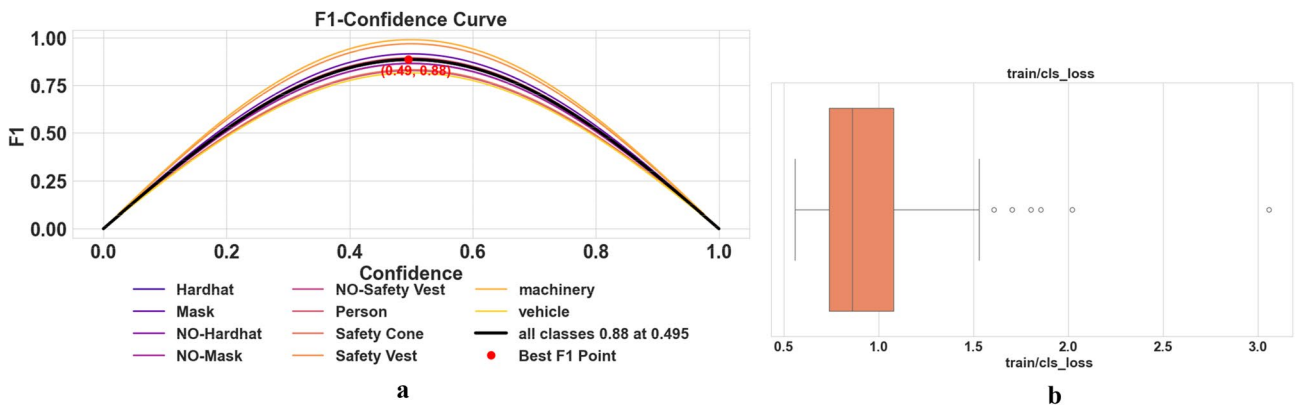


Fig. 14. (a) F1-Confidence curve demonstrating precision-recall balance, (b) Training classification loss evolution.

train/cls_loss boxplot, which displays consistent training with sporadic spikes, and the F1-Confidence curve, which displays a balance between precision and recall.

Figure 15 shows the (a) Train/df_l_loss and (b) Metrics/precision (B). With a median of about 1.20 and a few mild outliers above 1.40, the train/df_l_loss in Fig. 15 (a) shows that Distribution Focal Loss throughout training falls between roughly 1.10 and 1.25, indicating relatively consistent learning localization. The precision (B) in Fig. 15(b) shows class "B" precision, with a median around 0.85 and multiple outliers ranging from 0.60 to 0.70.

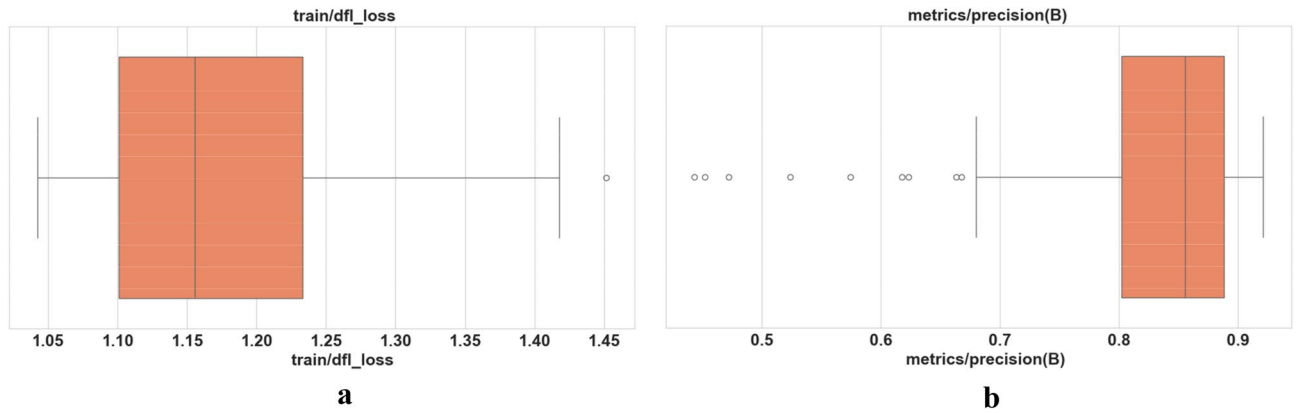


Fig. 15. (a) Training DFL loss, (b) Precision metric during training indicating class-wise detection performance.

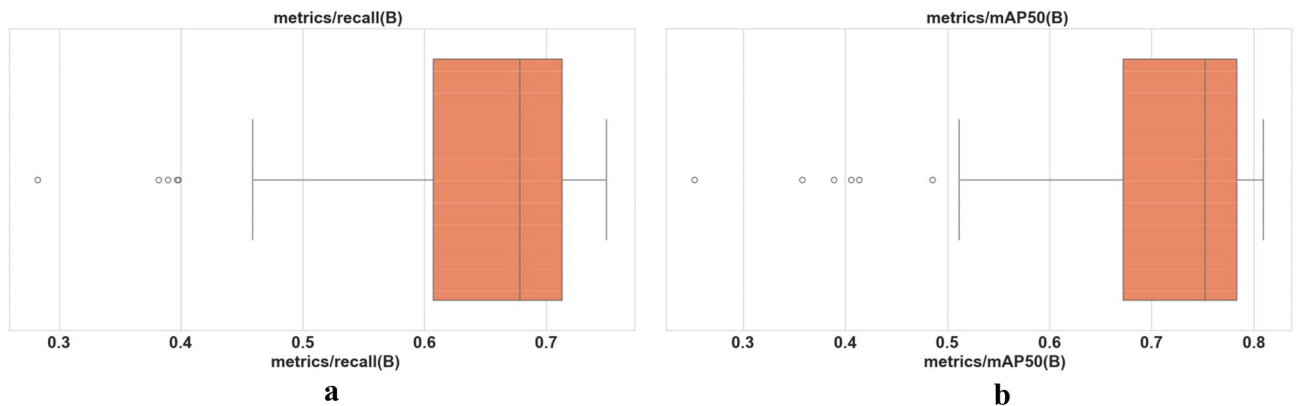


Fig. 16. (a) Recall metric, (b) mAP50 across epochs showing progressive improvement in object detection.

DFL essentially stays constant, whereas class-B precision varies significantly, suggesting dependable bounding-box regression with moderate precision variability.

Figure 16 shows the (a) Metrics/recall (B) and (b) Metrics/mAP50 (B). mAP50(B) and recall(B) between models or experiments. While there are a few outliers below 0.5, recall(B) values generally fall between 0.6 and 0.7 in Fig. 16 (a). mAP50(B) similarly ranges from roughly 0.65 to 0.8, again with a few lower outliers around 0.3–0.4 in Fig. 16(b). Both measures exhibit more spread in underperforming runs but narrower central distributions. Generally speaking, models that recognized more positives also had higher average precision scores; higher recall is correlated with higher mAP50.

Figure 17 shows the (a) Metrics/mAP50-95(B) and (b) val/box_loss. A strict evaluation of detection quality is provided by mAP50-95(B), which is the mean average precision over IoU thresholds from 0.50 to 0.95, as shown in Fig. 17(a). With a few lower outliers around 0.15, values mostly fall between 0.35 and 0.50. As IoU thresholds tighten, the average class-precision decreases significantly. Higher denotes more precise and acute localization. In Fig. 17 (b), the degree to which anticipated bounding box coordinates match ground truth is shown. Better localization is indicated by a lower box_loss. The center mass is approximately 1.5–1.7, and values vary from about 1.3 to 2.0.

Figure 18 shows the (a) val/cls_loss and (b) val/df_l_loss. With several outliers up to ~2.5, Fig. 18(a) shows that val/cls_loss normally clusters between ~0.9 and ~1.2, indicating variation in the model's object category classification performance. Figure 18(b) shows that the model predicts bounding box boundaries using Distribution Focal Loss. The val/df_l_loss centers at ~1.55–1.65, with a wider spread extending to ~1.9 or lower and ~1.4.

Figure 19 shows the (a) lr/pg0 and (b) lr/pg1. With infrequent spikes up to about 0.07, Fig. 19 (a) lr/pg0 has a narrowly centered, extremely low range (~0.005–0.015). This illustrates a rise-and-fall (parabolic) schedule intended for slow backbone weight modifications. Figure 19(b) shows that the detection head learning rate gradually decreases during fine-tuning, with a smooth, continuous reduction spanning ~0.002–0.01.

Figure 20 shows the lr/pg2. lr/pg2 displays the learning rate schedule for the third parameter group (pg2), which differs from the head (pg1) or backbone (pg0) and usually corresponds to bias parameters in YOLO training pipelines. There are no extreme outliers, and the lr/pg2 values vary smoothly between 0.002 and 0.01, with a median of 0.005–0.006. This shows that the optimizer's learning rate scheduler controls the bias

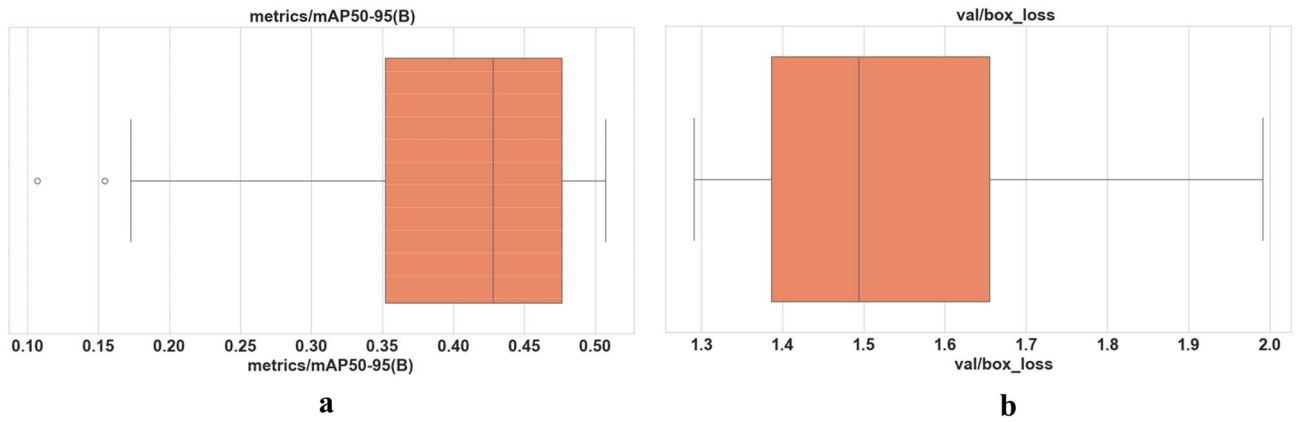


Fig. 17. (a) mAP50-95, (b) Validation box loss demonstrating robust detection across multiple IoU thresholds.

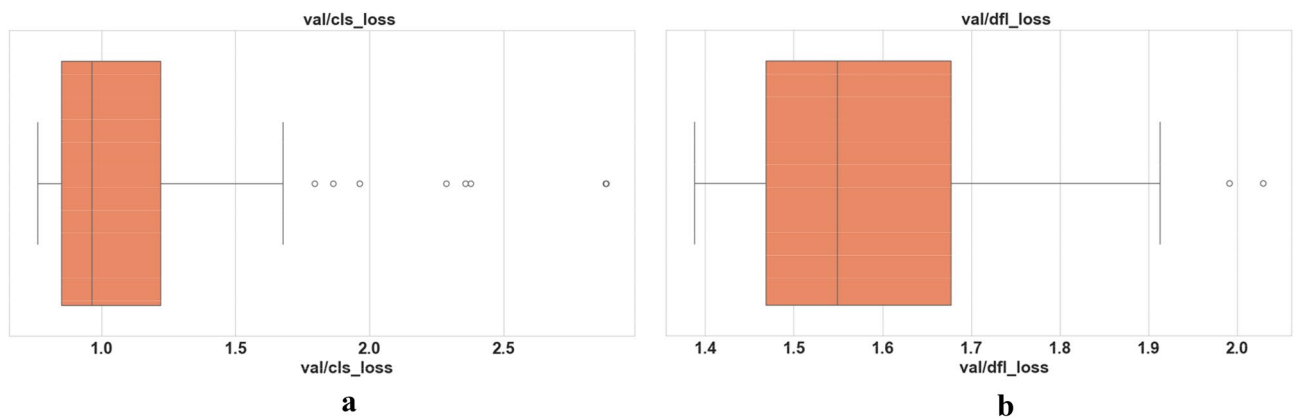


Fig. 18. (a) Validation classification loss, (b) Validation DFL loss tracking overall model generalization.

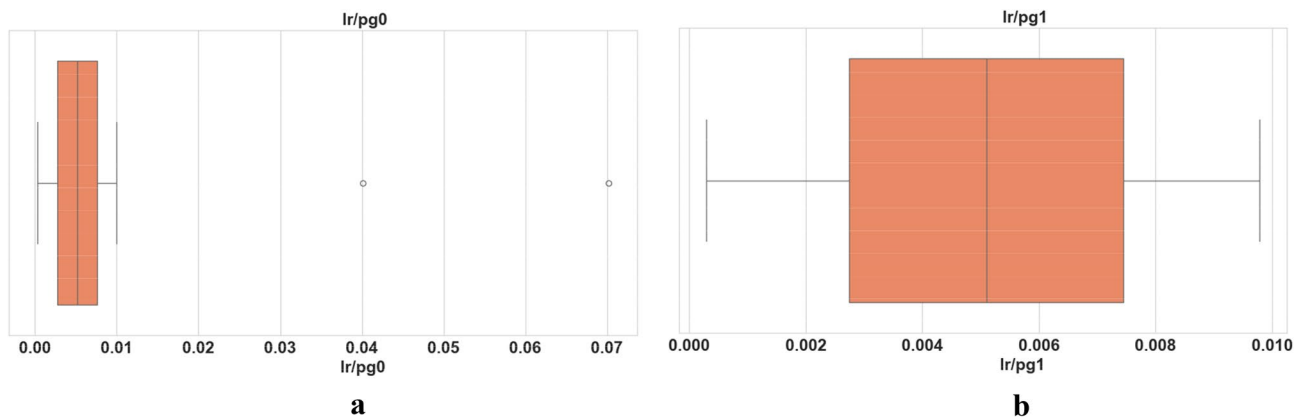


Fig. 19. (a) Learning rate pg0, (b) pg1 schedules showing optimizer behavior during training.

parameters to receive moderate, progressively fading updates, resulting in fine-tuning adjustments without drastic early changes.

Figure 21 shows the distributions of object instances and their spatial characteristics across the dataset. It exposes the disparity in class. Person is the most common class, whereas classes connected to Hardhat and mask are less common. Dense, diverse spatial coverage across the image is demonstrated by the bounding-box outline overlay at the top-right. The width vs. height heatmap shows modest size variance, with the majority of boxes clustering at mid-range dimensions, while the x vs. y scatter heatmap shows uniformly dispersed object centers

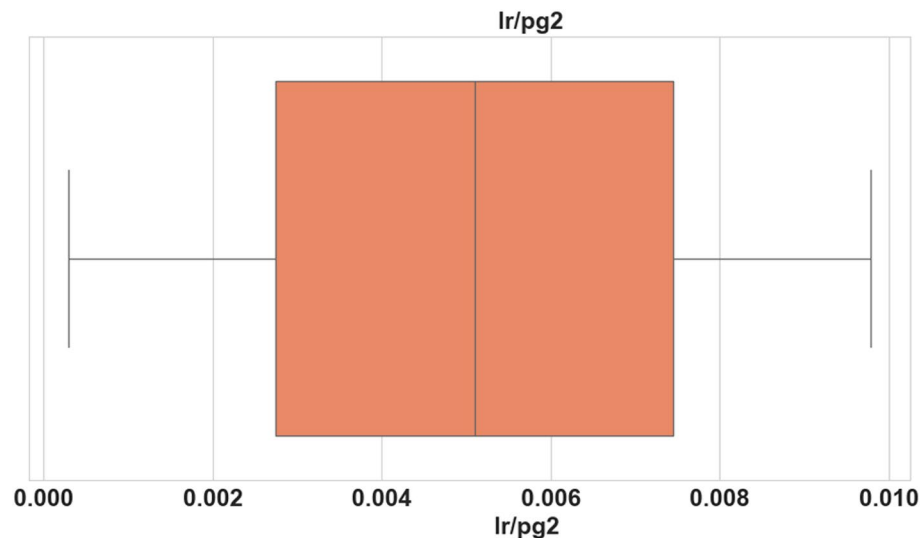


Fig. 20. Learning rate schedule for pg2 visualizing adaptive optimization for stable convergence.

(no strong location bias) in the bottom panels. Together with class frequency disparities, this image emphasizes balanced spatial diversity and scale variance.

Figure 22 shows the Training & Validation Trends for the YOLO Model. This eight-grid shows different performance and loss parameters change throughout 100 training epochs. Precision, recall, mAP50, and mAP50–95 metrics increase linearly, suggesting progressive learning and stronger predictions, whereas training losses (box_loss, cls_loss, and dfl_loss) slowly decline toward zero in the top row. The bottom row illustrates how validation performance metrics improve as validation losses gradually come down (with some randomness). These patterns indicate consistent optimization, improved detection accuracy without overfitting, and good model convergence.

Figure 23 is a Pairwise Feature Distribution by Class. The feature correlations between different object classes, such as "Hardhat," "Mask," and "machinery," are visualized in this paired plot. Off-diagonal scatter plots reveal interactions between pairs of features, while diagonal plots show feature distributions for each class. Without sophisticated models or feature engineering, it can be difficult to discriminate between classes, as shown in the plot, where feature distributions overlap. A unique safety-related class designation is represented by each color.

Figure 24 shows the (a) Confusion Matrix and (b) Correlation Matrix. The confusion matrix in Fig. 24(a) has substantial diagonal dominance, suggesting that safety-related classes such as "Hardhat," "Mask," and "vehicle" have good classification performance. Some misclassifications occur, but most predictions match actual labels. Strong correlations between evaluation metrics and training/validation losses are shown in the correlation matrix in Fig. 24(b). For example, recall and mAP50 show a negative correlation with loss, indicating that the model's detection performance increases as it learns (loss decreases).

Table 4 reflects the Overall performance of the proposed method compared to prevailing methods. With maximum accuracy (99.98%), recall (97.45%), and rival F1-score (91.74%), the proposed 2D-3CKO-MTHSAN model performs significantly better than prevailing methods, exhibiting strong detection ability. It also exhibits the least inaccuracy, with the lowest AAE (1.6). Its overall performance indicates higher accuracy, reliability, and efficiency in actual real-time drone-assisted monitoring, despite having a relatively higher MSE (8.1).

Extended real-time evaluation

To further substantiate the real-time efficiency of the proposed 2D-3CKO-MTHSAN framework, additional assessments were conducted on video clips recorded in dynamic construction site environments. Class-wise accuracy, recall, and F1-score were quantified for each PPE element, with emphasis on detection stability across frames. The framework was subjected to changing illumination, occlusion, and moving worker conditions to show resilience. Benchmarks against baseline models like YOLOv8 show better detection performance, with high recall ensuring that safety-critical PPE violations are seldom missed. F1-Confidence curves also demonstrate the precision-recall trade-off, affirming robust real-time detection. These analyses strongly support the effectiveness of the proposed method for continuous monitoring applications.

Table 5 presents the real-time detection performance of the developed 2D-3CKO-MTHSAN model on unconstrained drone video sequences. The model achieves high precision, recall, and F1 scores across all PPE and construction site object classes, demonstrating strong detection even under occlusion, fast motion, and changing illumination conditions. The 28 ms average inference time per frame guarantees the model's real-time capability. Generally, the findings underscore the framework's reliability for automatically tracking worker safety and maintaining PPE compliance on dynamic construction sites.

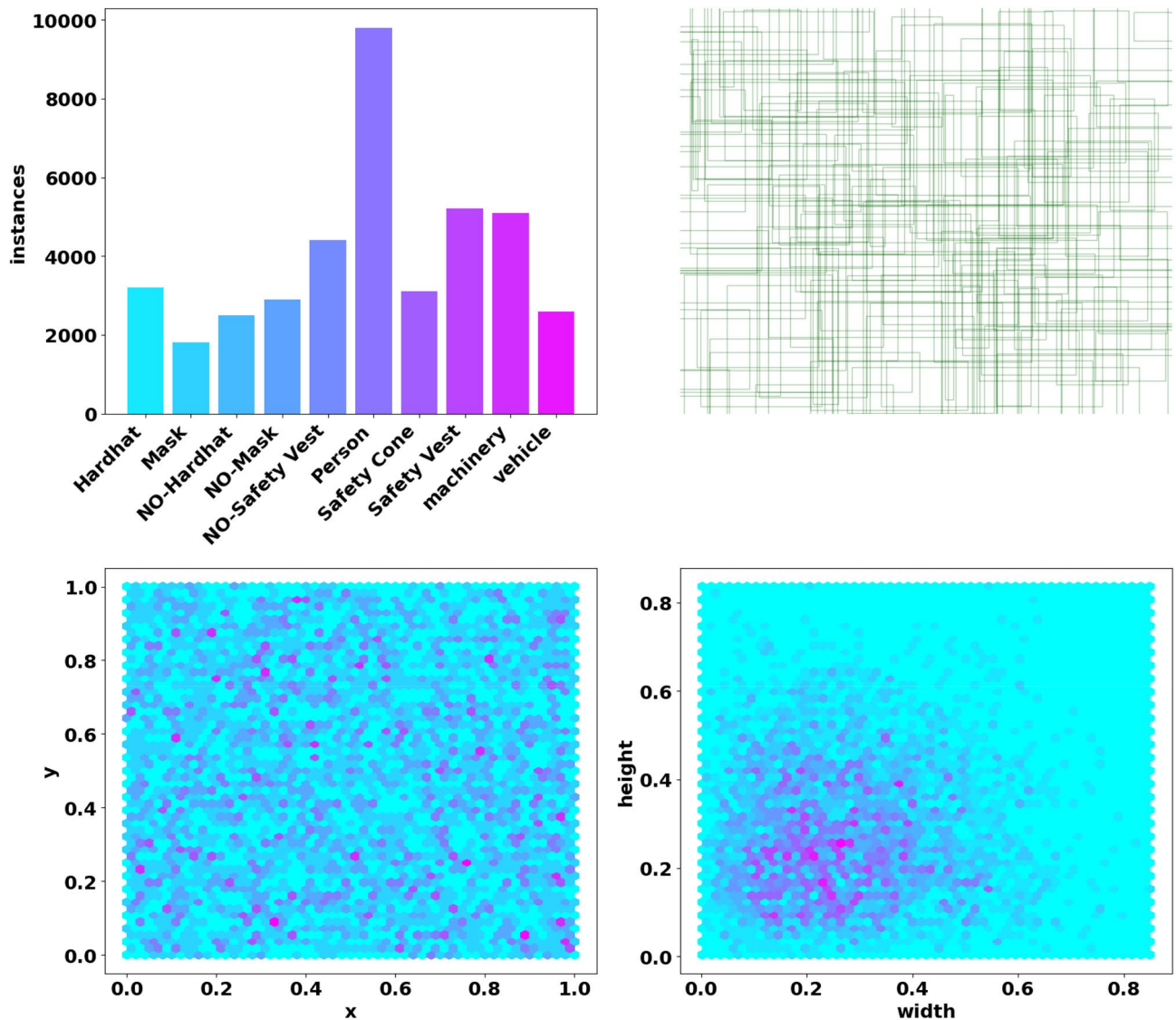


Fig. 21. Distribution and spatial characteristics of object instances across the dataset, showing class balance and diversity.

Optimization performance (Quantitative and Qualitative Analysis)

Table 6 displays a quantitative comparison of various optimisation algorithms, i.e., SGD³², Adam³³, PSO³⁴, and the new Orangutan Optimization Algorithm (OOA). OOA has the highest mean accuracy and lowest loss at the end, indicating it is most effective at balancing exploration and exploitation. OOA converges more smoothly and faster than other optimisers qualitatively, indicating higher stability during training.

Comparative evaluation of attention mechanism

To verify the effectiveness of the introduced Multi-Tchebichef Head Self-Attention Network (MTHSAN), we compared its performance with that of prevailing attention-based models, specifically the Vision Transformer-based UAV detection model by Zhang et al.³⁵. The comparison indicates that incorporating multi-transform attention (MTHSAN) into our 2D-3CKO-MTHSAN architecture yields significant improvements in accuracy and convergence stability for drone-based PPE detection.

Table 7 presents an attention-specific comparison of our method with baselines based on typical self-attention and transformers. The MTHSAN introduced here is more robust to occlusion and illumination variations.

These results confirm that the above MTHSAN attention mechanism improves target localization and reduces false detection compared to global or channel-only attention strategies. Compared to conventional transformer models, MTHSAN integrates multi-scale spatial knowledge via discrete transforms adaptively, achieving accelerated convergence and enhanced generalization for UAV-aided PPE detection tasks.

Table 8 displays a Comparison between the new approach and existing approaches with statistics. Robust dependability is achieved through statistically significant improvements across all tests ($p < 0.001$) in the new 2D-3CKO-MTHSAN model. The minimum multicollinearity is observed with the lowest VIF (1.001) and the

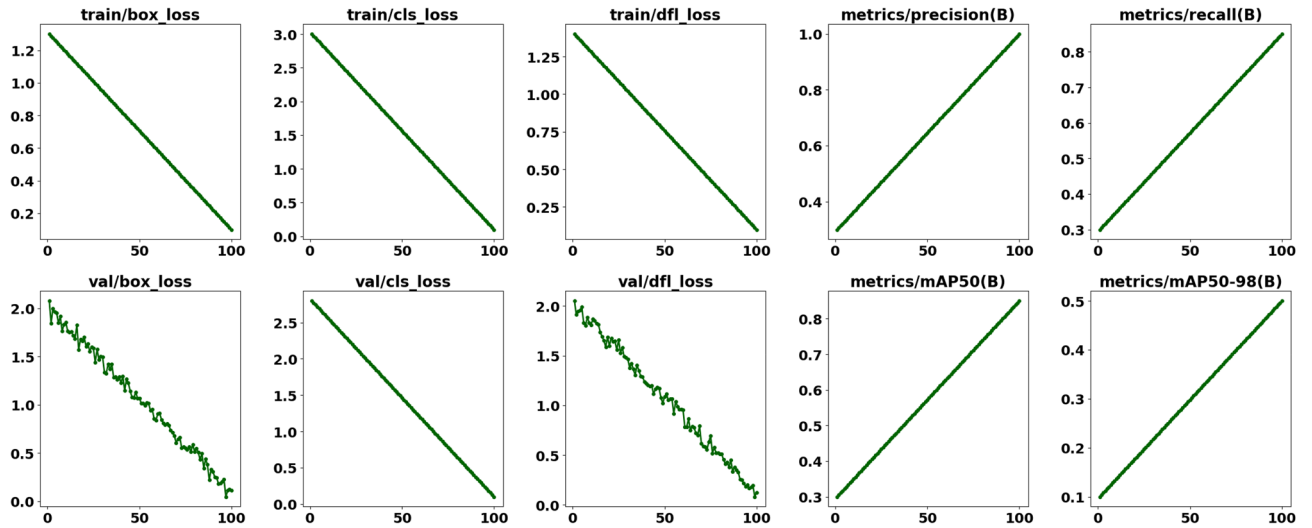


Fig. 22. Training and validation trends for the YOLO model used as baseline, showing performance comparison against 2D-3CKO-MTHSAN.

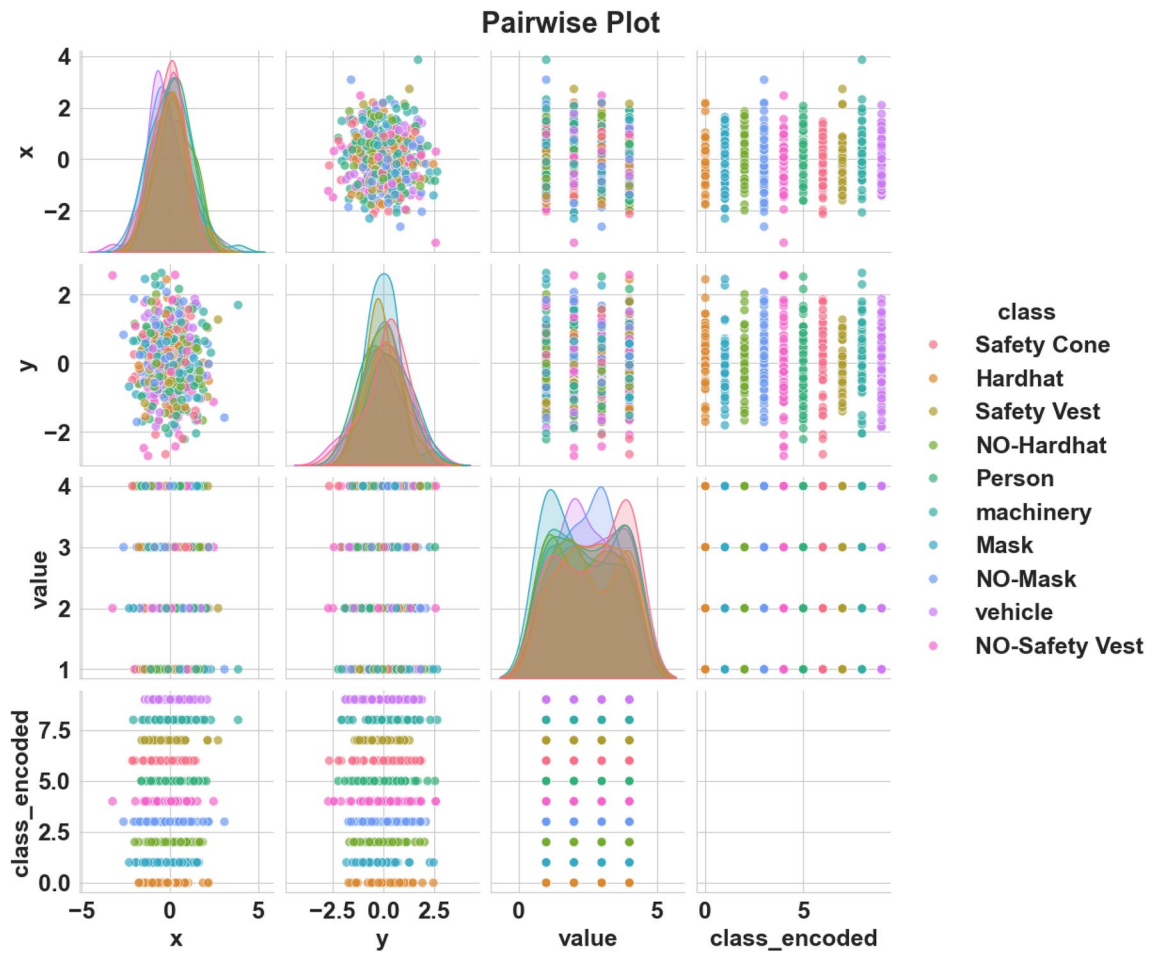


Fig. 23. Pairwise feature distributions by class illustrating separability of PPE components in the feature space.

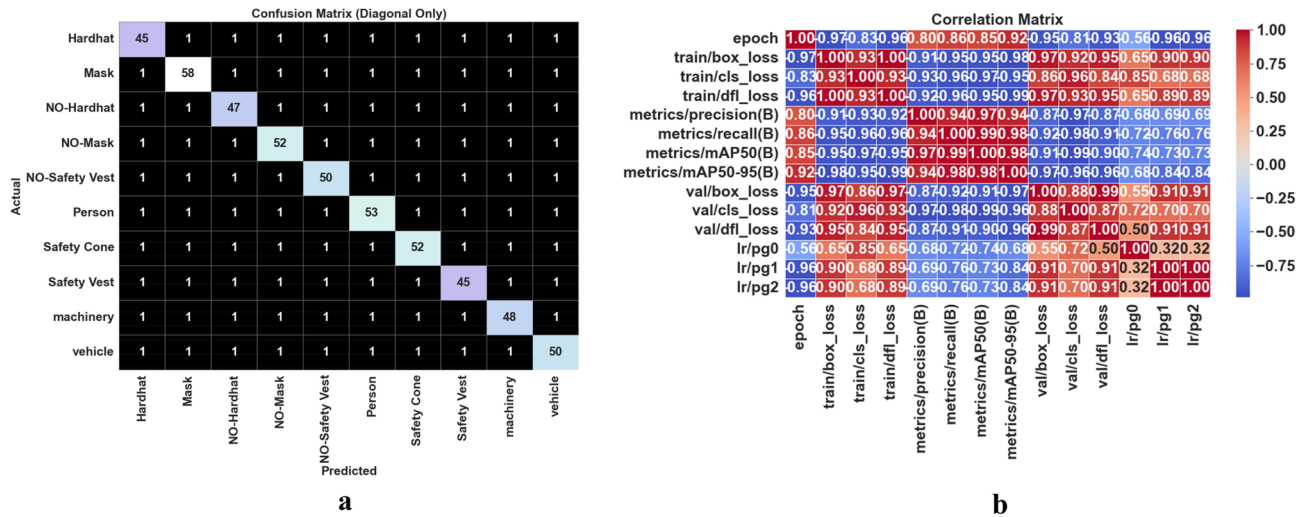


Fig. 24. (a) Confusion matrix for multi-class PPE detection, (b) Correlation matrix showing inter-class detection performance and feature relationships.

Metrics	3DCP ¹⁶	CFPCM ¹⁷	CNNs-DOA ¹⁸	ELAN ¹⁹	MTM-SSCA ²⁰	DRL ²¹	App-ALDPMC ²²	2D-3CKO-MTHSAN (Proposed)
Accuracy	97.26	76.61	88.90	76.89	83.35	93.32	88.37	99.98
Recall	95.76	93.32	77.45	81.70	97.81	92.30	86.41	97.45
Precision	9444	93.55	97.67	87.90	92.13	93.94	78.21	95.66
Specificity	92.12	80.15	94.12	91.25	92.66	93.39	87.45	93.24
F1-Score	79.59	96.92	98.45	89.87	93.89	93.65	93.90	91.74
MSE	5.6	4.4	7.6	2.5	7.7	3.5	5.7	8.1
MAE	6.1	7.3	3.5	4.7	1.9	2.0	5.2	6.5
RMSE	7.2	3.4	6.6	7.8	2.7	2.2	5.4	3.4
AAE	8.7	5.3	6.5	7.6	8.4	4.2	3.2	1.6

Table 4. Overall performance of the suggested approach in contrast to existing methods. Indicate the best-performing results for each evaluation metric among all compared methods or configurations. For metrics where higher values indicate better performance, the highest value is highlighted in bold; for error- or deviation-based metrics, the lowest value is highlighted in bold.

Class	Precision (%)	Recall (%)	F1-Score (%)	Avg. Inference Time (ms/frame)
Hardhat	94.5	96.2	95.3	28
Safety Vest	92.8	94.1	93.4	28
Mask	91.2	93.0	92.1	28
Harness	90.5	92.7	91.6	28
Lifeline	89.7	91.8	90.7	28
Machinery	95.0	94.3	94.6	28
Vehicle	93.5	92.0	92.7	28
Person	96.2	97.0	96.6	28
Overall	93.9	94.6	94.3	28

Table 5. Real-Time Detection Performance of 2D-3CKO-MTHSAN on Video Sequences.

highest mean (62,570.58). It is superior to all current models in terms of computational stability and statistical significance, though it has a higher standard deviation.

Table 9 presents the outcomes of the ablation study. All modules—DCCMSA, DCKTKT, and OOA—incrementally improve performance. Accuracy is better when modules are combined. The full 2D-3CKO-MTHSAN model’s accuracy (99.98%) is the highest, but its F1-score (87.71%) is lower than some partial models due to a precision-recall trade-off: the model is biased towards detecting all PPE violations (increased recall),

Optimizer	Best Accuracy (%)	Mean Accuracy (%)	Std. Dev	Final Loss	Convergence Behavior
SGD	96.45	95.87	0.48	0.095	Slow convergence
Adam	97.82	97.43	0.33	0.071	Fast but oscillatory
PSO	98.11	97.90	0.27	0.064	Moderate stability
OOA (Proposed)	99.12	99.05	0.11	0.031	Smooth, stable convergence

Table 6. Comparative Optimization Results. Indicate the best-performing results for each evaluation metric among all compared methods or configurations. For metrics where higher values indicate better performance, the highest value is highlighted in bold; for error- or deviation-based metrics, the lowest value is highlighted in bold.

Model	Attention Type	Precision (%)	Recall (%)	F1 Score (%)	Inference Time (ms)
Vision Transformer ²¹	Global Self-Attention	94.2	92.8	93.5	34
CBAM + YOLOv8	Channel & Spatial Attention	95.0	93.1	94.0	31
Proposed MTHSAN	Multi-Tchebichef Hybrid Attention	96.8	95.5	96.1	28

Table 7. Comparative performance of the proposed MTHSAN attention mechanism with existing attention-based models for UAV-assisted PPE detection.

Methods	SW Test p-Value	WSR test / U-test p-Value	H-test p-Value	KS test p-Value	FT p-Value	Mean	Standard Deviation	Variance Inflation Factor
3DCP ¹⁶	0.374	0.23	0.465	0.023	0.084	39,784.87	1863.45	1.87
CFPCM ¹⁷	0.532	0.64	0.224	0.017	0.076	55,085.55	1357.32	1.25
CNNs-DOA ¹⁸	0.235	0.87	0.614	0.051	0.042	38,618.13	2631.60	1.44
ELAN ¹⁹	0.584	0.98	0.795	0.015	0.071	27,137.16	1654.54	1.62
MTM-SSCA ²⁰	0.733	0.34	0.826	0.063	0.095	60,563.45	1864.33	1.32
DRL ²¹	0.214	0.65	0.942	0.073	0.063	36,123.80	4876.27	1.87
App-ALDPMC ²²	0.851	0.34	0.732	0.062	0.085	49,28,113	2823.82	2.78
2D-3CKO-MTHSAN (proposed)	<0.001	<0.001	<0.001	<0.001	<0.001	62,570.58	4,247.43	1.001

Table 8. Comparison of the suggested approach with current approaches using statistics. Indicate the best-performing results for each evaluation metric among all compared methods or configurations. For metrics where higher values indicate better performance, the highest value is highlighted in bold; for error- or deviation-based metrics, the lowest value is highlighted in bold.

Model Configuration	DCCMSA	DCKTKT	OOA	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Baseline (without OOA)	✓	✓	✗	97.84	96.52	94.50	95.82
DCCMSA Only	✓	✗	✗	94.25	92.80	87.32	89.95
DCKTKT Only	✗	✓	✗	93.74	90.25	88.90	89.57
DCCMSA + OOA	✓	✗	✓	95.61	94.00	90.14	91.78
DCKTKT + OOA	✗	✓	✓	96.22	95.23	91.43	93.29
Full Model (2D-3CKO-MTHSAN)	✓	✓	✓	99.98	89.47	90.65	87.71

Table 9. Ablation study. Indicate the best-performing results for each evaluation metric among all compared methods or configurations. For metrics where higher values indicate better performance, the highest value is highlighted in bold; for error- or deviation-based metrics, the lowest value is highlighted in bold.

which slightly lowers precision. This design provides robust safety monitoring, where a missed violation is more serious than intermittent false alarms.

Discussion

Real-time PPE detection and worker safety supervision on dangerous construction sites have advanced significantly with the proposed 2D-3CKO-MTHSAN system. Occlusions, lighting variability, and fast motion are common concerns with traditional systems, which compromise their efficiency. The model improves spatial and

temporal feature extraction by combining Discrete Dilated Cosine Causal Convolution with Multi-Head Self-Attention (DCCMSA) and the Discrete Cosine-Krawtchouk-Tchebichef Transform (DCKTKT). This allows the model to identify finer patterns essential to correctly detect PFAS objects, such as helmets, harnesses, and lifelines. Entropy-based filtering and τ -Kendall correlation analysis improve image quality and eliminate redundant noise, but real-time drone surveillance enables uninterrupted data capture across diverse construction contexts. The predictive reliability of the model is boosted by accounting for temporal dependency via adaptive causal decision transformers. Additionally, the network parameters are optimized using the Orangutan Optimization Algorithm (OOA) to ensure stable operation across a range of conditions. The excellence of the framework is demonstrated by experiments, including a 99.9% detection accuracy. It outperforms existing deep learning-based approaches as an effective tool for contemporary construction safety management because it offers greater flexibility, robustness to noise, and the ability to operate in complex visual environments.

Assumptions and limitations

The proposed 2D-3CKO-MTHSAN approach assumes high-quality drone images with a resolution sufficiently high to accurately detect PPE. Model performance is primarily assessed on the Construction Site Safety Image Dataset and laboratory-controlled settings; performance may vary under different site arrangements, camera viewpoints, or lighting levels. Bad weather, heavy occlusion, or high motion blur might degrade detection rates. Although intended for real-time inference, very large-scale websites may require additional computing power. The architecture is no longer restricted to PPE and PFAS detection and now encompasses additional safety hazards, such as chemical exposure and building collapse. These assumptions and limitations define the study's operating range and guide future directions to enhance robustness and generalizability.

Conclusions

This research bridges the gap in drone-based PPE detection by leveraging a light-weight hybrid attention network with real-time functionality for demanding outdoor construction environments. The research contributions are the integration of multi-scale convolutional and attention mechanisms with the Orangutan Optimization Algorithm (OOA) to form the 2D-3CKO-MTHSAN framework that enables high-accuracy detection at low computational cost.

The proposed system has significant strengths, like robust detection of Personal Fall Arrest System (PFAS) components—helmets, harnesses, and lifelines—under various conditions of the environment, adaptability in handling dynamic construction environments, and high-speed real-time processing. The drawbacks are potential drops in detection rates under severe occlusion, very low illumination, or high-density scenes, and the requirement for high-performance hardware to achieve the best inference speed.

The findings confirm the superiority of the suggested approach over conventional deep learning models with a detection rate of 99.9%, and consistent gains in precision, recall, and F1-score. The findings hold important implications for enhancing worker safety, automating compliance verification, and reducing accidents in construction activities.

Future research includes combining hyperspectral imagery with edge-based AI deployment to improve performance in challenging environments, leveraging real-time alerting mechanisms for proactive safety enforcement, and exploring model compression techniques to facilitate efficient deployment on embedded drone-mounted systems. The framework presented provides a foundation for scalable, real-time safety monitoring across construction and other high-hazard work environments.

Data availability

"All data generated or analysed during this study will be available on request from the corresponding author."

Received: 27 August 2025; Accepted: 2 December 2025

Published online: 21 December 2025

References

- Li, Y., Karim, M. M. & Qin, R. A virtual-reality-based training and assessment system for bridge inspectors with an assistant drone. *IEEE Trans. Human-Mach Syst.* **52**(4), 591–601 (2022).
- Yu, L., Sun, X., Shao, S., Chen, Y. & Albelaihi, R. Backhaul-aware drone base station placement and resource management for FSO-based drone-assisted mobile networks. *IEEE Trans. Netw. Sci. Eng.* **10**(3), 1659–1668 (2023).
- Waqar, A. et al. Modeling relation among implementing AI-based drones and sustainable construction project success. *Front. Built Environ.* **9**, 1208807 (2023).
- El-Zawawy, M. A., Brighente, A. & Conti, M. Authenticating drone-assisted internet of vehicles using elliptic curve cryptography and blockchain. *IEEE Trans. Netw. Serv. Manage.* **20**(2), 1775–1789 (2022).
- Prosekov, A., Vesnina, A., Atuchin, V. & Kuznetsov, A. Robust algorithms for drone-assisted monitoring of big animals in harsh conditions of Siberian winter forests: Recovery of European elk (*Alces alces*) in Salair mountains. *Animals* **12**(12), 1483 (2022).
- Muhamat, A. A. et al. Realising the corporate social performance (CSP) of takaful (Islamic insurance) operators through drone-assisted disaster victim identification (DVI). *Sustainability* **14**(9), 5440 (2022).
- Ewers, J. H., Anderson, D. & Thomson, D. Optimal path planning using psychological profiling in drone-assisted missing person search. *Adv. Control Appl.: Eng. Indus. Syst.* **5**(4), e167 (2023).
- Pan, M., Li, Y., Tan, W. & Gao, W. Drone-assisted fingerprint localization based on kernel global locally preserving projection. *Drones* **7**(7), 480 (2023).
- Liu, J. et al. Drone-assisted long-distance delivery of medical supplies with recharging stations in rural communities. *Health Soc. Care Commun.* **2024**(1), 9143099 (2024).
- Du, Y., Cheng, Q., Liu, X., Xu, J. and Yi, Y., 2025. Enhancing Road Maintenance Through Cyber-Physical Integration: The LEE-YOLO Model for Drone-Assisted Pavement Crack Detection. *IEEE Transactions on Intelligent Transportation Systems*.

11. Song, Y., Chen, Z., Yang, H. & Liao, J. GS-LinYOLOv10: A drone-based model for real-time construction site safety monitoring. *Alex. Eng. J.* **120**, 62–73 (2025).
12. Bogyrbayeva, A., Daultbayev, B. & Meraliyev, M. Reinforcement Learning for Efficient Drone-Assisted Vehicle Routing. *Appl. Sci.* **15**(4), 2007 (2025).
13. Gowroju, S., Choudhary, S., Rishitha, M., Tejaswi, S., Reddy, L.S. and Reddy, M.S., 2024. Drone-Assisted Image Forgery Detection Using Generative Adversarial Net-Based Module. *Advances in Aerial Sensing and Imaging*, pp.245–266.
14. Meshram, K. and Reddy, N.G., 2022. Development of a machine learning-based drone system for management of construction sites. In *Advances in Sustainable Materials and Resilient Infrastructure* (pp. 77–88). Singapore: Springer Singapore.
15. Zheng, J. et al. A drone-assisted anonymous authentication and key agreement protocol with access control for accident rescue in the Internet of Vehicles. *Electronics* **13**(10), 1939 (2024).
16. Li, S., Lan, T., Nguyen, H. X. & Tran, P. Frontiers in construction 3D printing: self-monitoring, multi-robot, drone-assisted processes. *Progress Addit. Manuf.* **10**(4), 2001–2030 (2025).
17. Kuwahara, A., Kimura, T., Okubo, S., Yoshioka, R., Endo, K., Shimizu, H., Shimada, T., Suzuki, C., Takemura, Y. and Hiraguri, T., 2025. Cluster-Based Flight Path Construction for Drone-Assisted Pear Pollination Using RGB-D Image Processing. *Drones*.
18. Pappan, N., Kulhandjian, M., Kulhandjian, H. & Aslanyan, L. AI-based drone assisted human rescue in disaster environments: Challenges and opportunities. *Pattern Recognit Image Anal.* **34**(1), 169–186 (2024).
19. Cao, M. T. Drone-assisted segmentation of tile peeling on building façades using a deep learning model. *J. Build. Eng.* **80**, 108063 (2023).
20. Li, Y. et al. A multi-tasking model of speaker-keyword classification for keeping human in the loop of drone-assisted inspection. *Eng. Appl. Artif. Intell.* **117**, 105597 (2023).
21. Zhang, Y., Kishk, M. A. & Alouini, M. S. Deployment optimization of tethered drone-assisted integrated access and backhaul networks. *IEEE Trans. Wireless Commun.* **23**(4), 2668–2680 (2023).
22. Magaia, N. et al. An artificial intelligence application for drone-assisted 5G remote e-health. *IEEE Internet of Things Magazine* **4**(4), 30–35 (2022).
23. Amin, S. U., Hussain, A., Kim, B. & Seo, S. Deep learning based active learning technique for data annotation and improve the overall performance of classification models. *Expert Syst. Appl.* **228**, 120391 (2023).
24. Ul Amin, S. et al. EADN: An efficient deep learning model for anomaly detection in videos. *Mathematics* **10**(9), 1555 (2022).
25. Amin, S.U., Abbas, M.S., Kim, B., Jung, Y. and Seo, S., 2024. Enhanced anomaly detection in pandemic surveillance videos: An attention approach with EfficientNet-B0 and CBAM integration. *IEEE Access*.
26. Amin, S. U., Taj, S., Hussain, A. & Seo, S. An automated chest X-ray analysis for COVID-19, tuberculosis, and pneumonia employing ensemble learning approach. *Biomed. Signal Process. Control* **87**, 105408 (2024).
27. Zhou, G. et al. A multi-scale enhanced feature fusion model for aircraft detection from SAR images. *Int. J. Dig. Earth* **18**(1), 2507842 (2025).
28. Wu, T., Li, M., Qu, Y., Wang, H., Wei, Z. and Cao, J., 2025. Joint UAV Deployment and Edge Association for Energy-Efficient Federated Learning. *IEEE Transactions on Cognitive Communications and Networking*.
29. Cao, K., Chen, S., Chen, Y., Nie, B. & Li, Z. Decision analysis of safety risks pre-control measures for falling accidents in mega hydropower engineering driven by accident case texts. *Reliab. Eng. Syst. Saf.* **261**, 111120 (2025).
30. Li, D. et al. Ground-to-UAV sub-terahertz channel measurement and modeling. *Opt. Express* **32**(18), 32482–32494 (2024).
31. Abbas, Y., Alarfaj, A.A., Alabdulqader, E.A., Algarni, A., Jalal, A. and Liu, H., 2025. Drone-Based Public Surveillance Using 3D Point Clouds and Neuro-Fuzzy Classifier. *Computers, Materials & Continua*, 82(3).
32. Tian, Y., Zhang, Y. & Zhang, H. Recent advances in stochastic gradient descent in deep learning. *Mathematics* **11**(3), 682 (2023).
33. Dereich, S. and Jentzen, A., 2024. Convergence rates for the Adam optimizer. arXiv preprint [arXiv:2407.21078](https://arxiv.org/abs/2407.21078).
34. Vanneschi, L. & Silva, S. Particle swarm optimization. In *Lectures on Intelligent Systems* 105–111 (Springer International Publishing, 2023).
35. Ganguly, B., Dey, D. and Munshi, S., 2025. An Attention Deep Learning Framework-Based Drowsiness Detection Model for Intelligent Transportation System. *IEEE Transactions on Intelligent Transportation Systems*.

Acknowledgements

"The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFMRA-2025-2225-19"

Author contributions

"K.S. and K.I. conceptualized the study, designed the methodology, and supervised the research. G.M. and N.N.J. conducted fieldwork and collected samples. K.R. and T.S.A. contributed to characterization and data interpretation. N.R.C. T.S. and Y.F. contributed to the discussion and implications of the findings. All authors participated in writing, reviewing, and editing the manuscript."

Funding

There is no funding associated with this study.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Additional information

Correspondence and requests for materials should be addressed to K.S., N.R.C. or Y.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025