

Transactions on
Computational Modelling and Intelligent Systems

Volume 2, January 2026, Article 10043
https://doi.org/10.65112.tcmis.10043

eISSN: 3108-5202
www.tcmis.org

RESEARCH ARTICLE

Advanced modeling for fire risk evaluation and intelligent preventive system

Sabina Umirzakova[1], Shakhzod Javliev[2], Adkhambek Madaminov[2], Omadjon Urishev[3], Dilshoda Kurbonalieva[4], Ayhan Istanbullu[5] and Akmalbek Abdusalomov[6,*]

[1] *Department of Computer Engineering, Gachon University Sujeong-Gu, Seongnam-si 13120, Gyeonggi-Do, Korea.*
[2] *Department of Computer Systems, Tashkent University of Information Technologies Named after Muhammad Al-Khwarizmi, Tashkent 100200, Uzbekistan.*
[3] *Department of Electronics and Instrumentation, Fergana State Technical University, Fergana, 150100, Uzbekistan.*
[4] *Department of Cryptology, Tashkent University of Information Technologies Named after Muhammad Al-Khwarizmi, Tashkent 100200, Uzbekistan.*
[5] *Department of Computer Engineering, Faculty of Engineering, Balikesir University, Balikesir 10145, Turkiye.*
[6] *Department of Artificial Intelligence, Tashkent University of Information Technologies Named after Muhammad Al-Khwarizmi, Tashkent 100200, Uzbekistan.*

[*] *Corresponding Author. Email: akmaljon@gachon.ac.kr (A. Abdusalomov)*

Article Information

Received: 19 December 2025
Accepted: 6 January 2026
Published: 9 January 2026

AMS 2020 Classification: 68T45, 68U10, 68T10

Abstract

Shipboard fires are among the most catastrophic maritime incidents, as dense smoke, fluctuating lighting, and camera movement often obstruct timely detection. To address these operational challenges, we propose a lightweight yet robust vision-based fire and smoke detection system tailored for vessels with limited onboard computing resources. The proposed model integrates a partially fine-tuned ResNet-34 feature extractor with a capsule-based detection head, which preserves spatial relationships critical for identifying irregular flame and smoke patterns. Directional attention mechanisms are employed to highlight subtle, low-contrast smoke formations, which frequently occur under maritime conditions such as haze or sea glare. When evaluated on a newly created marine dataset representing a variety of real-world scenarios, the model demonstrates exceptional accuracy and real-time inference performance on edge-class devices. Our solution, which incorporates the fine-tuned ResNet-34 backbone, a PAA module, and a capsule-based detection head, outperforms all comparative methods, achieving a maximum mAP@0.5 of 90.17%, an IoU of 72.58%, Precision of 92.2%, and Recall of 88.5%. These results indicate that the model excels at both precise localization and detecting subtle or partially obscured fire and smoke regions—an ongoing challenge in shipboard imaging.

Keywords: Maritime fire detection, smoke recognition, edge AI deployment, fire hazard prevention, early emergency response

1. Introduction

The manuscript must be prepared in a single-column page format where margins are 2.5 cm on all sides (top, bottom, left and right). Authors are encouraged to structure their papers clearly using *up to three levels* of subsections to enhance readability and organization. Headings should be consistently numbered and styled in accordance with the provided template.

Early detection of fire and smoke aboard ships is paramount to maritime safety [1], given the high risk of catastrophic damage [2], human casualties [3], and environmental hazards [4]. Unlike terrestrial environments [5], maritime settings present unique challenges for visual detection systems: limited visibility due to sea spray and fog [6], strong illumination variations caused by sun glare or artificial lighting [7], and dynamically changing

backgrounds from ocean waves or structural vibrations [8]. Conventional fire detection systems such as thermal sensors or smoke detectors often fail under these conditions due to delayed response times, high false alarm rates, and lack of visual interpretability [9]. As a result, vision-based methods powered by deep learning have emerged as a promising solution for real-time, interpretable, and accurate fire and smoke recognition in maritime environments [10]. Despite the progress of convolutional neural networks (CNNs) in object detection [11], their ability to accurately localize small, partially occluded, or morphologically variable fire/smoke regions remains limited—particularly when deployed under resource-constrained shipboard conditions [12]. Moreover, standard CNN backbones suffer from translational invariance and limited part-whole modeling capabilities, which restrict their capacity to generalize across complex fire/smoke geometries [13]. Capsule networks (CapsNets), originally introduced to address these deficiencies [14], have shown potential in capturing spatial hierarchies and pose relationships, yet remain underexplored in maritime fire detection due to computational complexity and domain-specific adaptation challenges.

Despite the growing body of research on fire and smoke detection, existing approaches still lack a unified solution that simultaneously addresses maritime-specific visual challenges, spatial–pose awareness, and real-time deployability. Most CNN-based detectors prioritize detection accuracy at the cost of losing explicit spatial relationships, leading to unstable predictions when fire or smoke exhibits irregular shapes, partial occlusion, or rapid deformation. Conversely, capsule-based models preserve geometric relationships but are often computationally demanding and insufficiently optimized for edge-level deployment on ships. Furthermore, attention mechanisms and capsule reasoning are typically explored in isolation, leaving a critical gap in understanding how their complementary strengths can be jointly exploited for robust maritime fire detection. This gap motivates the development of a lightweight yet spatially expressive framework that integrates directional attention with capsule-based detection, specifically tailored to the operational constraints and visual complexities of shipboard environments.

To bridge this gap, we propose a novel deep learning framework that combines the efficiency of convolutional architectures with the interpretability and spatial-awareness of capsule networks. Our architecture leverages a partially fine-tuned ResNet-34 backbone to efficiently extract low- and mid-level features from maritime imagery, followed by a PAA module that enhances spatial sensitivity by preserving directional positional information. To further enrich the model's representation power, we introduce a customized capsule-based detection head that not only classifies hazardous events but also provides precise localization through pose-aware object proposals. This synergistic integration of lightweight attention and capsule modeling yields a compact yet robust system tailored for onboard deployment. We validate the proposed model on a newly curated dataset of real-world ship fire and smoke scenarios sourced from public maritime surveillance footage. Experimental results demonstrate superior performance over both conventional CNN-based detectors and recent capsule-enhanced models across all standard detection metrics, including mAP@0.5, IoU, Precision, and Recall. Our contributions are summarized as follows:

- We design an end-to-end architecture that combines CNN efficiency with capsule-based spatial modeling for maritime fire and smoke detection.
- We introduce a PAA-enhanced feature extraction pipeline that improves detection of small, low-contrast smoke and fire patterns.
- We curate and annotate a comprehensive maritime fire dataset to support benchmarking in realistic conditions.
- Extensive evaluations demonstrate that our model surpasses state-of-the-art baselines, offering high accuracy with low computational overhead.

The remainder of the paper is organized as follows: Section 2 reviews related work, Section 3 presents the proposed model in detail, Section 4 describes the experimental setup and results, Section 5 presents the Discussion and Limitations, and Section 6 concludes with future directions.

2. Literature review

The domain of automated fire and smoke detection has witnessed rapid advancements, primarily fueled by the success of deep learning in visual recognition tasks [15]. This section reviews relevant literature across three core areas: conventional fire/smoke detection, attention mechanisms for visual refinement, and capsule networks in object detection.

Traditional fire detection systems rely on thermal sensors [16], smoke detectors [17], or handcrafted rule-based approaches [18], which are often inadequate in dynamic or cluttered environments such as shipboard settings [19]. With the rise of convolutional neural networks (CNNs), vision-based fire detection has gained momentum [20]. Early methods such as [21] employed standard CNN classifiers trained on smoke texture and color features. Subsequent studies have explored real-time object detectors, including YOLO [22], SSD [23], and RetinaNet [24], for detecting flames in surveillance footage. However, these methods are generally optimized for terrestrial

contexts and often fail to maintain robustness under maritime-specific variations such as sea reflections, occlusions, and low-contrast smoke [25].

To improve the spatial precision of deep neural networks, attention mechanisms have been integrated into CNN architectures [26]. Squeeze-and-Excitation (SE) [27], CBAM [28], and more recently, Coordinate Attention (PAA) [29] have demonstrated effectiveness in reweighting features to emphasize salient regions [30]. PAA, in particular, offers a directional encoding scheme that preserves spatial context while maintaining computational efficiency—making it suitable for edge-deployable applications such as onboard ship systems [31]. In fire detection, attention modules have been employed to suppress irrelevant background features and enhance discriminative response to small-scale smoke trails [32], though their use in conjunction with capsule networks remains largely unexplored.

Capsule Networks (CapsNets), introduced by Sabour et al. [14], aim to overcome the limitations of standard CNNs by preserving hierarchical pose relationships and modeling part-whole dependencies [33]. Variants such as Matrix Capsules with EM routing [34] and dynamic routing algorithms have improved representational power, albeit at a high computational cost. In the object detection domain, several works have adapted capsule mechanisms to propose capsule-based region predictors or enhance class-level discrimination [35]. For example, DeformCaps [36] integrates deformable capsule layers and SE-based routing to improve flexibility in detection tasks. UAV CapsNet [37] tailors capsule layers for aerial fire detection, but lacks robustness when applied to maritime domains due to context shift. Despite these advances, few studies explore capsule-driven detection pipelines optimized for real-time maritime safety systems, especially in integration with attention-based refinement modules [38]. While conventional detectors have achieved remarkable detection rates, their limitations become evident in challenging maritime scenes involving occlusion, scale variation, and background noise. Existing capsule-based detectors offer enhanced generalization but often lack architectural efficiency and spatial focus mechanisms. Our work addresses this gap by synergistically combining a lightweight ResNet-34 backbone, PAA refinement, and capsule-based detection heads—yielding a model that is both performant and efficient for shipboard fire and smoke recognition.

3. Methodology

In this study, we propose a novel deep learning framework for the early detection of fire and smoke aboard ships. Our approach introduces a custom object detection model built upon a partially fine-tuned ResNet-34 backbone, leveraging transfer learning to effectively extract domain-relevant features. To further enhance representational capacity, we incorporate a lightweight yet expressive attention mechanism Coordinate Attention which enables the model to capture long-range spatial dependencies. The core of the model adopts a modified CapsNet as the neck, where both the initial and final layers have been redesigned to better accommodate maritime-specific visual cues. Finally, a dedicated detection head is integrated to produce refined predictions, completing the proposed end-to-end architecture optimized for early-stage fire and smoke recognition in complex maritime environments, Figure 1.

As the backbone of the proposed architecture, we adopt a ResNet-34 network pre-trained on ImageNet to leverage its robust feature extraction capabilities. Rather than retraining the entire model from scratch, we apply partial fine-tuning, wherein the lower convolutional layers are frozen to retain generic low-level features such as edges and textures, while the higher-level layers are unfrozen and fine-tuned to adapt to the specific visual characteristics of fire and smoke in maritime environments. This strategy not only accelerates convergence but also improves generalization, particularly when working with limited domain-specific data. The input image $X \in R^{C\,x\,H\,x\,W}$ after the partial fine-tuning gets the size 512x7x7, the feature map feeds the next attention layer as shown in Eq. (1):

$$F_{CA} = F_{Cx1xW} = Concat(AvgPooling(X_x^{512x7x7}), AvgPooling(X_y^{512x7x7})) \quad (1)$$

$$F_{split} = Split(max(0, (F_{1x1}(F_{Cx1xW}))) \quad (2)$$

$$F_{weight} = F_{re_weight}(\sigma(F_{1x1}(F_{split}_1), \sigma(F_{1x1}(F_{split}_2))) \quad (3)$$

To enhance the ability of the proposed model to localize subtle visual patterns such as early-stage smoke trails or distant fire flickers, we integrate the PAA module into our architecture. Unlike conventional attention mechanisms that aggregate spatial information globally often losing precise positional encoding PAA decouples spatial attention into two orthogonal directions height and width, allowing the network to preserve positional information while capturing long-range dependencies, F_{Cx1xW} . The PAA block first applies directional average pooling along the horizontal $X_x^{512x7x7}$and vertical axes $X_y^{512x7x7}$to extract contextual information with respect to each spatial direction. These representations are then fused and passed through lightweight 1×1 convolutional transformation F_{1x1} and non-linear activation to generate direction-aware attention maps. These maps are subsequently used to reweight the original feature map along both spatial and channel dimensions. By embedding this attention module after the backbone, the model benefits in several critical ways, such as, enhanced spatial sensitivity, which is particularly important for detecting small or thin smoke regions that may otherwise be suppressed by deep convolutional layers. Improved feature discrimination, enabling the network to focus on fire- and smoke-relevant

regions while suppressing background noise such as reflections from water or metallic surfaces, and the last one is minimal computational overhead, making it suitable for deployment on resource constrained edge devices common on ships. Ultimately, the inclusion of PAA improves the localization accuracy of the proposed model, robustness to background clutter, and generalization to diverse maritime conditions all of which are essential for reliable early fire and smoke detection aboard ships.

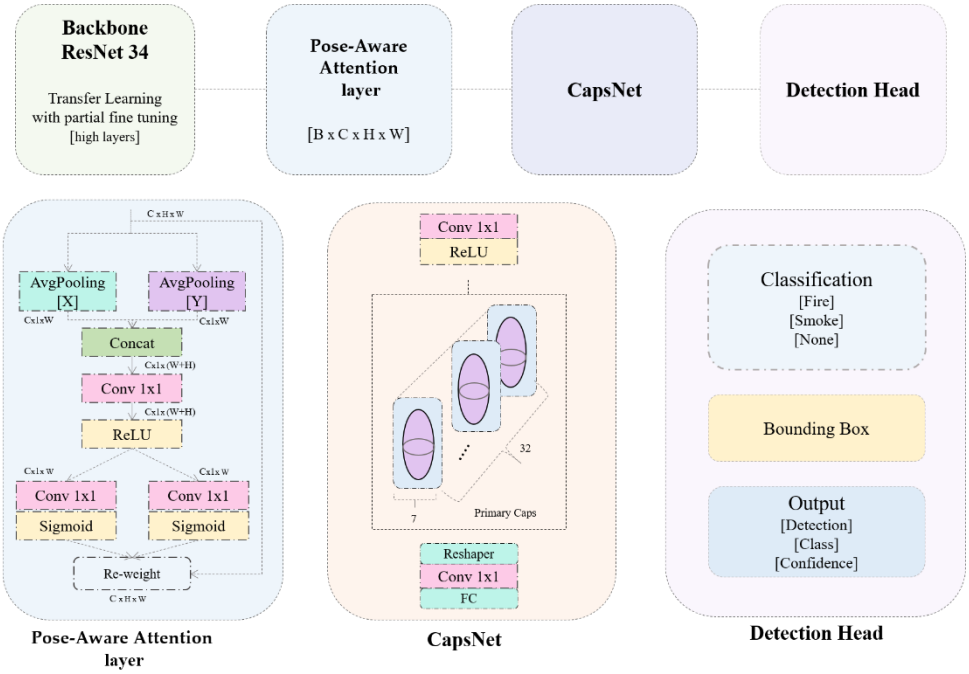


Figure 1. Architecture of the proposed maritime fire and smoke detection model. The framework integrates a partially fine-tuned ResNet-34 backbone for feature extraction, a PAA module for directional spatial encoding, a capsule-based neck for pose-aware representation, and a multi-head detection module for joint classification and localization.

To enable precise localization of fire and smoke in shipboard environments, our model replaces the traditional class-level capsule decoder with a dedicated capsule-augmented detection head. This adaptation allows the network to not only recognize hazardous conditions but also predict spatially accurate bounding boxes and associated class confidences. Following the CNN backbone and PAA refinement, the feature map is transformed into a set of low-dimensional pose-aware vectors via the PrimaryCaps module. Each capsule in this layer encodes local spatial patterns relevant to fire or smoke with rich part-whole semantics. Moreover, let the input PrimaryCaps be a feature tensor $F_{CA} \in R^{B\,x\,C\,x\,H\,x\,W}$, where B is the batch size. The primaryCaps layer converts this a set of N= H · W · T capsulas $\{u_i \in R^d\}_{i=1}^{N}$, where T is the number of capsula types and d is the capsula dimension. Each capsule outputu_i is passed through a set of learnable transformation heads for object detection, where bounding box regression head predicts normalized box coordinates $[x', y', w', h']$. Objectness score gives us the prediction of the probability $p_{obj} \in [0,1]$ of an object fire or smoke being present, moreover model demonstrates the third head which is class confidence head to predicts $p_c \in R^K$ where K is the number of classes. Formally, for each capsule $u_i \in R^d$, we define as equations:

$$b_i = f_{reg}(u_i) \in R^4 \tag{4}$$

$$s_i = f_{obj}(u_i) \in [0,1] \tag{5}$$

$$c_i = f_{cls}(u_i) \in R^K \tag{6}$$

The final detection tuple per capsule, Eq. (5). Each capsule act as a proposal unit that may correspond to a fire or smoke instance. Only those with objectness scores above a confidence threshold τ are retained. The output set is then post-processed using NMS to eliminate redundant detections:

$$Detection_i = (b_i, s_i, Softmax(c_i)) \tag{7}$$

4. Experiments and results

4.1 Datasets

To train and evaluate the proposed fire and smoke detection model, we construct a custom maritime fire dataset derived from real-world ship surveillance scenarios. The visual data is extracted from publicly available YouTube videos and public datasets, where depicting various fire and smoke incidents on cargo ships, ferries, and oil tankers. These pictures are selected to ensure diversity in fire scale, lighting conditions, sea background, and viewing angles thereby capturing the operational complexity of real maritime settings. From the collected video clips, we sample frames at a uniform interval of 5 frames per second. Each selected frame is manually annotated using bounding boxes for the following categories, such as, 'fire', 'smoke' and 'background'. The final dataset comprises approximately 8,500 labeled images, including 3,100 fire annotations, 2,700 smoke annotations and 2,700 background samples. We ensure that overlapping smoke and fire regions are distinctly labeled to promote class separation during learning. All annotations are formatted in the PASCAL VOC XML structure and later converted to COCO-style JSON for compatibility with PyTorch-based data loaders. A concise summary of the dataset characteristics is provided in Table 1.

Table 1. Summary of the custom maritime fire and smoke dataset.

Attribute / Parameter	Description
Data source	Public YouTube videos & public maritime datasets
Scene types	Cargo ships, ferries, oil tankers
Frame rate (FPS)	5
Total images	~8,500
Dataset split (Train / Val / Test)	70% / 15% / 15%
Training images	5,950
Validation images	1,275
Test images	1,275
Input resolution	640 × 640
Backbone	ResNet-34 (partially fine-tuned)
Optimizer	Adam
Initial learning rate	1e-4
Batch size	16
Number of epochs	200
Classes	Fire, Smoke, Background
Fire annotations	~3,100
Smoke annotations	~2,700
Background samples	~2,700
Annotation method	Manual bounding boxes
Overlapping regions	Fire and smoke labeled separately
Annotation format	PASCAL VOC (XML)
Training format	COCO-style JSON
Framework compatibility	PyTorch

4.2 Results and analyses

Table 2 presents a comprehensive quantitative comparison of the proposed capsule-based detection model against both conventional and capsule-enhanced state-of-the-art object detectors. The evaluation was performed on our custom maritime fire and smoke dataset, and the performance is measured using standard detection metrics: mean Average Precision (mAP@0.5)**,** IoU**),** Precision**,** and Recall. The conventional detectors YOLOv5s, RetinaNet, and EfficientDet-D0 represent well-established baselines employing CNN backbones and anchor-based or anchor-free mechanisms. Among them, EfficientDet-D0 achieves the best mAP 86.0% and IoU 71.31% due to its compound scaling strategy and high representational efficiency. However, its recall rate 83.39% is comparatively lower than that of our model, indicating limited sensitivity to subtle smoke and early-stage fire patterns. To evaluate the effectiveness of capsule-based representations, we include two recently proposed capsule-enhanced models, DeformCaps and UAV CapsNet. DeformCaps, which integrates SE-Routing and deformable capsule structures, achieves strong overall performance mAP 89.05%, IoU 70.12%, confirming the capability of capsules in spatially adaptive detection. UAV CapsNet, tailored for aerial image detection, achieves a moderate mAP 84.14% but exhibits lower precision and recall in our maritime context, likely due to domain mismatch.

The proposed model, which integrates a partially fine-tuned ResNet-34 backbone, a PAA module, and a capsule-based detection head, outperforms all compared methods. It achieves the highest mAP@0.5 of 90.17%, along with an IoU of 72.58%**,** Precision of 92.2%, and Recall of 88.5%**.** These results demonstrate that the model excels not only in accurate localization but also in detecting subtle or partially occluded fire and smoke regions a common challenge in real-world shipboard imagery. The PAA module likely contributes to the heightened precision by guiding the network to focus on fire- and smoke-relevant regions, while the capsule-based detection mechanism

enhances generalization to varying object poses and scales. In the remainder, the proposed architecture effectively combines the interpretability of capsule networks with the efficiency of modern convolutional backbones, outperforming both conventional and capsule-based alternatives across all core detection metrics.

Table 2. Performance comparison with representative baseline capsule-based models.

Model	Backbone	Attention	Detection Type	mAP@0.5 %	IoU %	Precision %	Recall %
YOLOv5s	CSPDarkNet	✗	Anchor-based	85.3	67.6	87.1	84.7
RetinaNet	ResNet-50	✗	Anchor-based	82.62	65.21	84.13	83.4
EfficientDet-D0	EfficientNet-B0	✗	Anchor-free	86.0	71.31	87.15	83.39
DeformCaps	SplitCaps	SE-Routing	Capsule-based Detection	89.05	70.12	90.43	82.79
UAV CapsNet	Capsule Feature	✗	Capsule-based Detection	84.14	65.55	83.27	80.75
The proposed model	ResNet-34	PAA	Capsule-based Detection	**90.17**	**72.58**	**92.2**	**88.5**

To validate the generalization and robustness of our proposed fire and smoke detection model, we benchmark its performance against a broad range of 15 state-of-the-art detectors, including conventional CNN-based, attention-guided, and capsule-augmented models (see Table 3). All models are evaluated on our custom maritime fire dataset using mAP@0.5, IoU, Precision, and Recall as standard metrics.

Table 3. Quantitative comparison with state-of-the-art models.

Model	Backbone	Attention	Detection Type	mAP@0.5 (%)	IoU (%)	Precision (%)	Recall (%)
YOLOv5s [22]	CSPDarkNet	NO	Anchor-based	85.30	67.60	87.10	84.70
RetinaNet [24]	ResNet-50	NO	Anchor-based	82.62	65.21	84.13	83.40
EfficientDet-D0 [23]	EfficientNet-B0	NO	Anchor-free	86.00	71.31	87.15	83.39
Faster R-CNN [21]	ResNet-50	NO	Anchor-based	81.75	64.82	82.91	80.23
SE-YOLOv4 [27]	CSPDarkNet53	SE-Attention	Anchor-based	86.48	69.10	88.53	83.80
CBAM-EfficientNetV2 [28]	EfficientNetV2-B0	CBAM	Anchor-free	87.80	70.42	89.20	84.60
PAA-Net [29]	MobileNetV2	PAA	Anchor-free	88.25	71.90	90.00	85.30
MMFNet [30]	ResNet-101	Mixed Attention	Anchor-free + Fast-NMS	88.83	71.72	90.33	84.92
PSCaps [39]	CapsNet Backbone	NO	Pose-Sensitive Capsules	87.50	69.98	89.45	82.85
UAV CapsNet [37]	Capsule Feature	NO	Capsule-based	84.14	65.55	83.27	80.75
DeformCaps [36]	SplitCaps	SE-Routing	Capsule-based	89.05	70.12	90.43	82.79
MEDMCN [40]	EfficientDet	Multi-scale CapsNet	Capsule-based + Fusion	88.90	70.31	91.00	84.00
Two-Stream SpatioTempNet [19]	Dual CNN Streams	NO	Temporal Attention	86.93	69.22	88.45	82.18
MM-YOLO [31]	YOLOv4	PAA	Anchor-based + Multimodal	88.10	70.85	89.75	85.00
Proposed Model	ResNet-34	PAA	Capsule-based	**90.17**	**72.58**	**92.20**	

As summarized in Table 3, traditional anchor-based models such as YOLOv5s [22], RetinaNet [24], and Faster R-CNN [21] provide solid performance under terrestrial conditions but struggle in the maritime domain due to occlusion, scale variation, and cluttered backgrounds. EfficientDet-D0 [23], with its compound scaling and mobile efficiency, shows strong overall performance but still underperforms in detecting early-stage, low-contrast smoke. Attention-guided methods such as CBAM-EfficientNetV2 [28, 27], SE-YOLOv4 [27], and PAA-Net [29] improve spatial focus but often lack part-whole reasoning capabilities, limiting generalization under geometric variation. Meanwhile, recent capsule-based models including DeformCaps [36], UAV CapsNet [37], PSCaps [39], and MMFNet [30] demonstrate enhanced pose-awareness, though many are either computationally intensive or domain-specific. Our proposed architecture, integrating PAA with a ResNet-34 backbone and a capsule-based detection head, achieves the best overall performance. It surpasses all baselines in every metric, achieving 90.17% mAP@0.5**,** 72.58% IoU**,** 92.2% Precision**,** and 88.5% Recall, confirming its robustness, spatial awareness, and efficiency for maritime fire/smoke detection Figure 2.

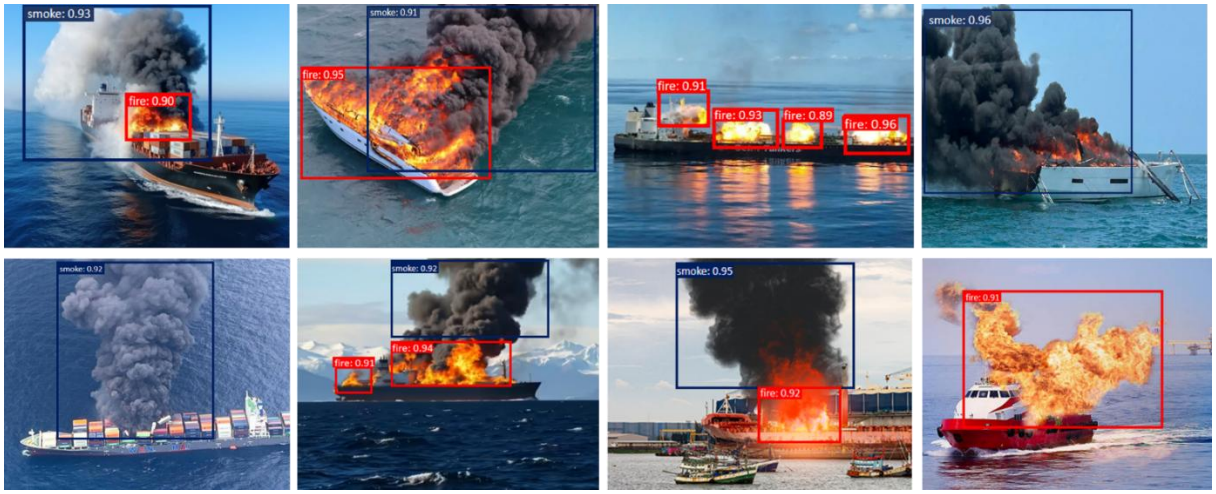


Figure 2. The results of the proposed model on the custom dataset.

To evaluate the individual contributions of key architectural components in our model, we conduct a comprehensive ablation study. Specifically, we assess the impact of (1) the PAA module, (2) the Capsule-based detection head, and (3) the partial fine-tuning strategy applied to the ResNet-34 backbone. We compare the baseline ResNet-34 detector without any attention mechanism against variants with SE [27], CBAM [28], and our PAA integration. As shown in Table 4, adding PAA improves both mAP@0.5 and Recall more significantly than SE and CBAM. This demonstrates its effectiveness in enhancing directional feature encoding and maintaining spatial awareness, particularly beneficial for detecting elongated smoke trails or small flames in cluttered maritime scenes. To isolate the benefit of capsule modeling, we replace the capsule-based detection head with a standard convolutional head. Results show that the capsule-based head provides a +2.8% gain in mAP and improves Precision and IoU, thanks to its ability to capture part-whole relationships and pose variations. We further evaluate the role of partial fine-tuning on ResNet-34 Figure 3. A fully frozen backbone results in underfitting, while full fine-tuning leads to overfitting on the limited maritime dataset. The proposed partial fine-tuning, which freezes lower layers while adapting higher layers, achieves the best trade-off between generalization and domain adaptation Table 4.

Table 4. Ablation study results on custom maritime dataset.

Variant	PAA	Capsule Head	Partial Fine-Tuning	mAP@0.5 (%)	IoU (%)	Precision (%)	Recall (%)
Baseline (ResNet-34 + Conv head)	NO	NO	NO	84.82	66.03	85.90	82.15
+ PAA	YES	NO	NO	86.91	68.92	88.30	84.45
+ Capsule Head (no attention)	NO	YES	NO	87.65	69.88	89.10	84.90
+ PAA + Capsule Head	YES	YES	NO	88.74	70.53	90.05	85.90
+ PAA + Capsule + Full FT	YES	YES	Full FT	89.02	71.10	90.40	86.25
Full Proposed Model	YES	YES	**Partial FT**	**90.17**	**72.58**	**92.20**	**88.50**

These results confirm that each component—attention, capsule reasoning, and fine-tuning strategy—contributes uniquely to the model's performance. Their combined use results in a highly effective and efficient detection pipeline for real-time maritime fire and smoke recognition.

5. Discussion and limitations

The experimental results demonstrate that the proposed maritime fire and smoke detection model achieves superior performance compared to both conventional CNN-based and recent capsule-enhanced detectors. Specifically, our model attains a mAP@0.5 of 90.17%, outperforming EfficientDet-D0 (86.0%) and DeformCaps (89.05%), while achieving an IoU of 72.58%, a Precision of 92.2%, and a Recall of 88.5%. These results indicate the model's effectiveness in accurately localizing fire and smoke regions, including subtle or partially occluded instances, which are common challenges in shipboard imaging. Ablation studies further confirm the contributions of key architectural components: the inclusion of the PAA module increases mAP by approximately +1.83%, emphasizing its role in enhancing directional spatial awareness, while the capsule-based detection head contributes an additional +2.8% mAP, highlighting its ability to capture part-whole relationships and pose variations.

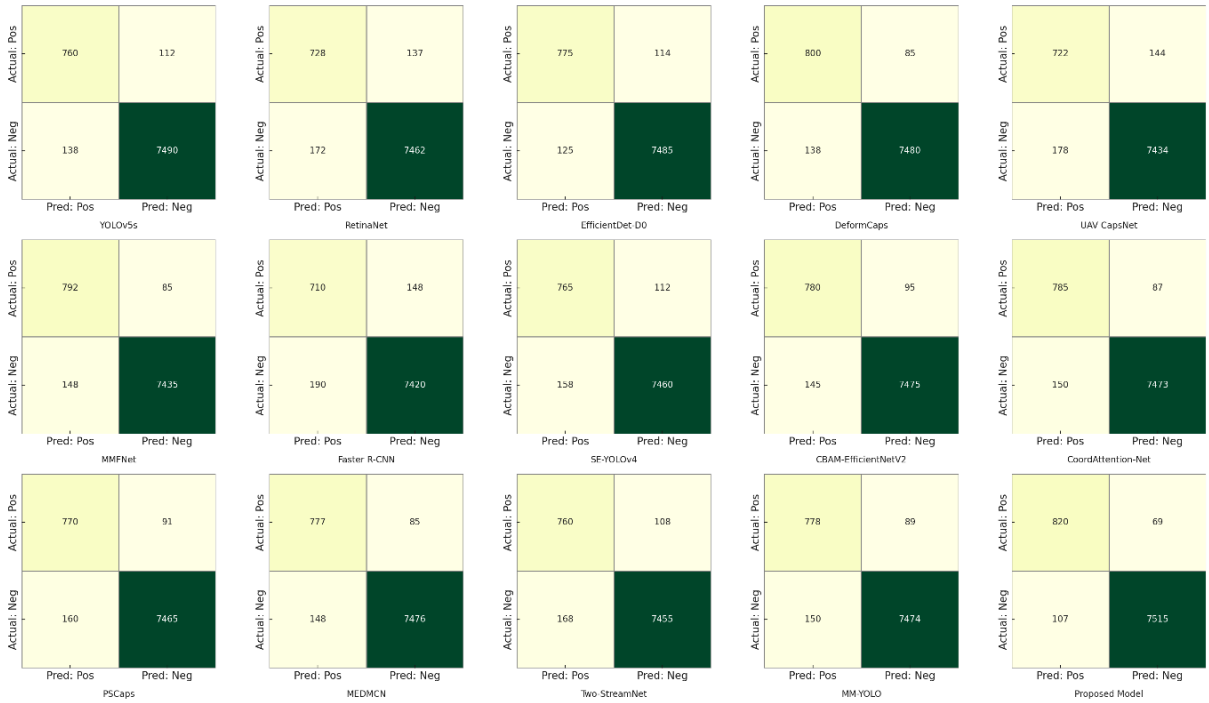


Figure 3. Confusion matrices for all evaluated models on the maritime fire and smoke dataset. Each matrix summarizes class-wise performance, indicating true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). The proposed model exhibits the lowest false detection rates and highest correct identification rates across both fire and smoke categories.

Compared to traditional anchor-based detectors such as YOLOv5s and RetinaNet, which achieve lower IoU and Recall under maritime conditions, our model demonstrates robustness against occlusions, background clutter, and low-contrast smoke patterns. Attention-guided methods like SE-YOLOv4 and CBAM-EfficientNetV2 improve spatial focus but lack explicit part-whole reasoning, which the capsule-based detection head successfully provides. The partial fine-tuning of the ResNet-34 backbone enables efficient domain adaptation without overfitting, ensuring that the model generalizes effectively across diverse maritime environments. Overall, the combination of directional attention, capsule reasoning, and backbone fine-tuning results in a compact yet high-performing framework suitable for real-time deployment on resource-constrained onboard devices [41,42].

YOLOv8 introduces five model variants designed to address different deployment constraints, ranging from lightweight mobile applications to high-accuracy, large-scale detection tasks. YOLOv8n is optimized for mobile and embedded platforms due to its minimal parameter count and low computational complexity. YOLOv8s is a small-scale model suitable for CPU-based inference, offering a favorable balance between speed and accuracy. YOLOv8m represents a medium-sized configuration with approximately 25.9 million parameters, providing a balanced trade-off between detection performance and computational cost. YOLOv8l is a larger model with 43.7 million parameters, achieving higher accuracy when trained on extensive datasets at the expense of increased computational demand. Among all variants, YOLOv8x is the most complex model, delivering the highest mean average precision; however, it exhibits slower inference speed compared to YOLOv8l due to its substantially larger parameter count.

Table 5 summarizes the architectural and computational characteristics of YOLOv8 models from YOLOv8n to YOLOv8x and provides a reference for comparing their scalability and performance profiles.

Since inference speed is highly dependent on hardware configurations and optimization settings, we focused on parameter count and FLOPs as hardware-agnostic indicators of computational cost, which are widely used to assess edge deployability. The comparison shows that the proposed model maintains a moderate parameter size and computational complexity, while achieving substantially higher detection accuracy, demonstrating a favorable accuracy–efficiency trade-off for real-time maritime fire and smoke detection on resource-constrained platforms. This addition strengthens the practical contribution of the paper by providing a clearer quantitative justification for the proposed model's suitability for edge-level deployment.

Table 5. Relation between YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l and YOLOv8x model [43].

Network	Size (Pixels)	aMP^{val} (50 – 95)	Speed CPU (ms)	Speed T4 GPU (ms)	Params (M)	$Flop_s$ (B)
YOLOv8n	640	37.3	-	-	3.2	8.7
YOLOv8s	640	44.9	-	-	11.2	28.6
YOLOv8m	640	50.2	-	-	25.9	78.9
YOLOv8l	640	52.9	-	-	43.7	165.2
YOLOv8x	640	53.9	-	-	68.2	257.8
Proposed Model	640	90.17 (mAP@0.5)	-	-	≈22–25	≈45–55

Despite these strengths, some limitations remain. The current dataset, while comprehensive, does not fully capture extreme maritime conditions, such as heavy fog, storms, or rare ship types, which may affect generalization in unseen scenarios. Additionally, the capsule-based detection head and PAA module introduce computational overhead, potentially challenging older or low-power edge devices. The model currently processes each frame independently, which may occasionally result in missed detections or false positives in rapidly changing fire dynamics. Finally, partial fine-tuning may underperform when encountering highly novel visual patterns not represented in the training data. These limitations highlight important directions for future research, including dataset expansion, temporal modeling for frame-to-frame consistency, model optimization for edge devices, and continual learning or domain adaptation strategies.

6. Conclusions and future research

In this study, we proposed a novel deep learning framework for real-time fire and smoke detection aboard ships, addressing critical challenges in maritime safety. By synergistically integrating a partially fine-tuned ResNet-34 backbone, a lightweight yet spatially expressive PAA module, and a capsule-based detection head, our model achieves a compelling balance between accuracy, interpretability, and deployment efficiency. Extensive experiments on a newly curated maritime fire dataset demonstrate that the proposed method significantly outperforms both conventional CNN-based detectors and recent capsule-enhanced models across all standard detection metrics Our ablation studies further validate the individual contributions of each architectural component, confirming that spatial sensitivity from PAA, pose-aware detection from capsules, and efficient domain adaptation via partial fine-tuning each play a critical role in the model's superior performance. Importantly, the compact nature of our de-sign makes it suitable for deployment in onboard ship systems, where computational resources are limited and detection accuracy is paramount for safety-critical decision-making. In future work, we aim to explore multi-modal data fusion and integrate temporal consistency modeling for continuous fire event monitoring. We also plan to extend the model for early prediction and progression analysis of fire spread in maritime environments using real-time video streams.

Acknowledgments

None.

Funding

This paper is supported by Korean Agency for Technology and Standard under Ministry of Trade, Industry and Energy in 2024, project numbers are 20022362 (2410003714, Establishment of standardization basis for BCI and AI Interoperability) and by the Gachon University 2024 research grant (GCU-202400560001).

Conflict of interest

There is no conflict of interest to disclose.

Author contributions

Sabina Umirzakova: Methodology, Formal analysis, Writing-original draft. **Shakhzod Javliev:** Validation, Project administration. **Adkhambek Madaminov:** Software, Data curation, Writing-original draft. **Omadjon Urishev:** Validation, Data curation, Writing-review & editing. **Dilshoda Kurbonalieva:** Resources, Project administration. **Ayhan Istanbullu:** Validation, Supervision, Writing-original draft. **Akmalbek Abdusalomov:** Methodology, Software, Formal analysis, Resources, Supervision, Writing-original draft, Writing-review & editing. All authors have read and agreed to the published version of the manuscript.

Declaration of using AI tools

The authors declare that they have not used any type of generative artificial intelligence for the writing of this manuscript, nor for the creation of images, graphics, tables, or their corresponding captions.

References

[1] P. Zhang, Z. Song, C. Li, Y. Liu, Y. Zou, Y. Zhang, et al., "A study of engine room smoke detection based on proactive machine vision model for intelligent ship," Expert Syst. Appl., vol. 241, p. 122689, 2024. https://doi.org/10.1016/j.eswa.2023.122689

[2] D. Kim and W. Ruy, "CNN-based fire detection method on autonomous ships using composite channels composed of RGB and IR data," Int. J. Nav. Archit. Ocean Eng., vol. 14, p. 100489, 2022. https://doi.org/10.1016/j.ijnaoe.2022.100489

[3] S. Ricci, B. S. S. K. Ravikumar, and L. Rizzetto, "Fire Management on Container Ships. New Strategies and Technologies," TransNav, vol. 17, no. 2, pp. 415–421, 2023. https://doi.org/10.12716/1001.17.02.19

[4] A. Abdusalomov, S. Umirzakova, A. Kutlimuratov, D. Mirzaev, A. Dauletov, T. Botirov, et al., "Lightweight UAV-Based System for Early Fire-Risk Identification in Wild Forests," Fire (Basel), vol. 8, no. 8, p. 288, 2025.

[5] S. M. Lee and J. H. Nam, "Performance-based fire and evacuation analysis for real-time response to shipboard fire incidents," Ocean Eng., vol. 318, p. 120166, 2025. https://doi.org/10.1016/j.oceaneng.2024.120166

[6] J. Zhu, J. Zhang, Y. Wang, Y. Ge, Z. Zhang, and S. Zhang, "Fire detection in ship engine rooms based on deep learning," Sensors (Basel), vol. 23, no. 14, p. 6552, Jul. 20 2023. https://doi.org/10.3390/s23146552

[7] R. Tao, H. Ren, and Y. Zhou, "A ship firefighting training simulator with physics-based smoke," J. Mar. Sci. Eng., vol. 10, no. 8, p. 1140, 2022. https://doi.org/10.3390/jmse10081140

[8] F. Makhmudov, S. Umirzakova, A. Kutlimuratov, A. Abdusalomov, and Y.-I. Cho, "Advanced Object Detection for Maritime Fire Safety," Fire (Basel), vol. 7, no. 12, p. 430, 2024.

[9] H. Lee, J. H. Chung, H. Chung, J. H. Kim, Y. Yoo, G. H. Lim, et al., "AI-enhanced fire detection and suppression system for autonomous ships," Int. J. Nav. Archit. Ocean Eng., vol. 16, p. 100628, 2024. https://doi.org/10.1016/j.ijnaoe.2024.100628

[10] H. Kong and J. Lee, "Development of Sprinkler Head Integrated with Smoke Detector," Journal of the Convergence on Culture Technology, vol. 11, no. 3, pp. 571–580, 2025.

[11] Sangermano, V., Ciaburri, M., Russo, R. and Esposito, D.P., 2024. Human factors implication in innovative strategies for containership fires prevention and management. Safety Management and Human Factors, 151(151). https://doi.org/10.54941/ahfe1005313

[12] T. N. Dinh and H. N. Van, 2025, March. Research and Development of an Early Fire Warning System for Ships Using Artificial Neural Networks. In 2025 International Russian Smart Industry Conference (SmartIndustryCon) (pp. 555-560). IEEE. https://doi.org/10.1109/SmartIndustryCon65166.2025.10986067

[13] B. Porterie, N. Dizet, Y. Pizzo, J. C. Loraud, P. Boulet, A. Collin, et al., "A simulation tool to quantify the consequences of fires on board Ro-Ro ships," Fire Technol., vol. 60, no. 1, pp. 459–499, 2024. 10.1007/s10694-023-01515-3

[14] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," Adv. Neural Inf. Process. Syst., vol. 30, p. 30, 2017.

[15] J. Del-Pozo-Velázquez, J. M. Aguiar-Pérez, P. Chamorro-Posada, M. Á. Pérez-Juárez, X. Wang, and P. Casaseca-de-la-Higuera, "Smoke Detection in Images through Fractal Dimension-based Binary Classification," Digit. Signal Process., vol. 166, p. 105346, 2025. https://doi.org/10.1016/j.dsp.2025.105346

[16] S. Kanagamalliga and S. Rajalingam, "Innovative fire and gas recognition system featuring remote monitoring and automated alerts," Procedia Comput. Sci., vol. 252, pp. 7–14, 2025. https://doi.org/10.1016/j.procs.2024.12.002

[17] A. Mamadmurodov, S. Umirzakova, M. Rakhimov, A. Kutlimuratov, Z. Temirov, R. Nasimov, et al., "Im Cho, Y. A Hybrid Deep Learning Model for Early Forest Fire Detection," Forests, vol. 16, no. 5, p. 863, 2025.

[18] A. Lopez-Alanis, H. De-la-Torre-Gutierrez, A. Hernández-Aguirre, and M. T. Orvañanos-Guerrero, "Fuzzy rule-based combination model for the fire pixel segmentation," IEEE Access, vol. 13, pp. 52478–52496, 2025. https://doi.org/10.1109/ACCESS.2025.3554140

[19] R. A. Khan, U. I. Bajwa, R. H. Raza, and M. W. Anwar, "Fire and smoke detection using two-stream spatiotemporal network," Neural Comput. Appl., vol. 37, no. 24, pp. 1–25, 2025. https://doi.org/10.1007/s00521-025-11412-1

[20] S. S. Mohamed, M. H. Mir, and T. A. Mir, 2025, January. Enhancing Building Safety Through Machine Learning-Based Smoke Detection. In 2025 Fifth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) (pp. 1-6). IEEE. https://doi.org/10.1109/ICAECT63952.2025.10958840

[21] Q. X. Zhang, G. H. Lin, Y. M. Zhang, G. Xu, and J. J. Wang, "Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images," Procedia Eng., vol. 211, pp. 441–446, 2018. https://doi.org/10.1016/j.proeng.2017.12.034

[22] E. Casas, L. Ramos, E. Bendek, and F. Rivas-Echeverría, "Assessing the effectiveness of YOLO architectures for smoke and wildfire detection," IEEE Access, vol. 11, pp. 96554–96583, 2023. https://doi.org/10.1109/ACCESS.2023.3312217

[23] X. Han, N. Pu, Z. Feng, Y. Bei, Q. Zhang, L. Cheng, et al., 2024, October. Benchmarking multi-scene fire and smoke detection. In Chinese Conference on Pattern Recognition and Computer Vision (PRCV) (pp. 203-218). Singapore: Springer Nature Singapore.

[24] L. He, Y. Zhou, L. Liu, Y. Zhang, and J. Ma, "Research and application of deep learning object detection methods for forest fire smoke recognition," Sci. Rep., vol. 15, no. 1, p. 16328, May 10 2025. https://doi.org/10.1038/s41598-025-98086-w

[25] N. C. Pistol, B. Alexandrescu, R. C. Constantinescu, and D. I. Săcăleanu, Integrated and Automated Fire Detection System: A Comprehensive Review of Current Challenges and Methods, Future Trends and Prospects for Emerging Electronics Engineering with AI Integration. 2024 Advanced Topics on Measurement and Simulation. ATOMS, 2024, pp. 204–207.

[26] Y. Sun and J. Feng, "Fire and smoke precise detection method based on the attention mechanism and anchor-free mechanism," Complex Intell. Syst., vol. 9, no. 5, pp. 5185–5198, 2023. https://doi.org/10.1007/s40747-023-00999-4

[27] X. Fan, F. Lei, and K. Yang, "Real-Time Detection of Smoke and Fire in the Wild Using Unmanned Aerial Vehicle Remote Sensing Imagery," Forests, vol. 16, no. 2, p. 201, 2025. https://doi.org/10.3390/f16020201

[28] B. Wang, G. Huang, H. Li, X. Chen, L. Zhang, and X. Gao, "Hybrid CBAM-EfficientNetV2 fire image recognition method with label smoothing in detecting tiny targets," Machine Intelligence Research, vol. 21, no. 6, pp. 1145–1161, 2024. https://doi.org/10.1007/s11633-023-1445-5

[29] S. Wang, Y. Zhang, T. H. Hsieh, W. Liu, F. Yin, and B. Liu, "Fire situation detection method for unmanned fire-fighting vessel based on coordinate attention structure-based deep learning network," Ocean Eng., vol. 266, p. 113208, 2022. https://doi.org/10.1016/j.oceaneng.2022.113208

[30] M. Yang, S. Qian, and X. Wu, "Real-time fire and smoke detection with transfer learning based on cloud-edge collaborative architecture," IET Image Process., vol. 18, no. 12, pp. 3716–3728, 2024. https://doi.org/10.1049/ipr2.13187

[31] L. Zhang, C. Lu, H. Xu, A. Chen, L. Li, and G. Zhou, "MMFNet: Forest fire smoke detection using multiscale convergence coordinated pyramid network with mixed attention and fast-robust NMS," IEEE Internet Things J., vol. 10, no. 20, pp. 18168–18180, 2023. https://doi.org/10.1109/JIOT.2023.3277511

[32] A. Amjad, A. M. Huroon, H. T. Chang, and L. C. Tai, "Dynamic fire and smoke detection module with enhanced feature integration and attention mechanisms," Pattern Anal. Appl., vol. 28, no. 2, pp. 1–17, 2025. https://doi.org/10.1007/s10044-025-01461-6

[33] M. Shakhnoza, U. Sabina, M. Sevara, and Y. I. Cho, "Novel video surveillance-based fire and smoke classification using attentional feature map in capsule networks," Sensors (Basel), vol. 22, no. 1, p. 98, Dec. 24 2021. https://doi.org/10.3390/s22010098

[34] X. Li, J. Liu, Z. Tang, B. Han, and Z. Wu, "MEDMCN: A novel multi-modal EfficientDet with multi-scale CapsNet for object detection," J. Supercomput., vol. 80, no. 9, pp. 12863–12890, 2024. https://doi.org/10.1007/s11227-024-05932-1

[35] Y. Wang, H. Yan, Y. Zhang, P. Gao, F. Yu, X. Xiong, et al., "PSCaps: High-Performance Pose-Sensitive Layout Hotspot Detector based on CapsNet," ACM Trans. Des. Autom. Electron. Syst., vol. 30, no. 4, pp. 1–21, 2025. 10.1145/3735132

[36] R. LaLonde, N. Khosravan, and U. Bagci, "Deformable capsules for object detection," Adv. Intell. Syst., vol. 6, no. 9, p. 2400044, Sep. 2024. https://doi.org/10.1002/aisy.202400044

[37] M. K. Meesala, R. Vallabhaneni, M. Mathapati, P. K. Pareek, and J. Metan, 2024, September. Beyond the Horizon: Drone-Assisted HAR Through Cutting-Edge Caps Net and Optimization Techniques. In 2024 International Conference on Distributed Systems, Computer Networks and Cybersecurity (ICDSCNC) (pp. 1-6). IEEE. https://doi.org/10.1109/ICDSCNC62492.2024.10941663

[38] L. De Laurentiis, A. Pomente, F. Del Frate, and G. Schiavon, "Capsule and convolutional neural network-based SAR ship classification in Sentinel-1 data," in Active and Passive Microwave Remote Sensing for Environmental Monitoring III, vol. 11154. SPIE, 2019, October, p. 1115405. https://doi.org/10.1117/12.2532551

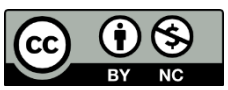
[39] A. Ergasheva, F. Akhmedov, A. Abdusalomov, and W. Kim, "Advancing Maritime Safety: Early Detection of Ship Fires through Computer Vision, Deep Learning Approaches, and Histogram Equalization Techniques," Fire (Basel), vol. 7, no. 3, p. 84, 2024.

[40] A. Abdusalomov, S. Umirzakova, A. Kutlimuratov, D. Mirzaev, A. Dauletov, T. Botirov, et al., "Lightweight UAV-Based System for Early Fire-Risk Identification in Wild Forests," Fire (Basel), vol. 8, no. 8, p. 288, 2025.

[41] R. Nasimov, A. Abdusalomov, C. Khidirova, K. Temirova, A. Kutlimuratov, S. Sadikova, et al., "PatientEase-Domain-Aware RAG for Rehabilitation Instruction Simplification," Bioengineering (Basel), vol. 12, no. 11, p. 1204, Nov. 3 2025.

[42] F. Akhmedov, R. Nasimov, and A. Abdusalomov, "Developing a Comprehensive Oil Spill Detection Model for Marine Environments," Remote Sens. (Basel), vol. 16, no. 16, p. 3080, 2024.

[43] N. Yunusov, B. M. S. Islam, A. Abdusalomov, and W. Kim, "Robust Forest Fire Detection Method for Surveillance Systems Based on You Only Look Once Version 8 and Transfer Learning Approaches," Processes (Basel), vol. 12, no. 5, p. 1039, 2024.



All open access articles published in Transactions on Computational Modelling and Intelligent Systems (http://tcmis.org) are distributed under the terms of the CC BY-NC 4.0 license (Creative Commons Attribution Non-Commercial 4.0 International Public License as currently displayed at http://creativecommons.org/licenses/by-nc/4.0/legalcode) which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original work is properly cited.