

An Integrated Lightweight UAV Vision and Relation-Aware Language Intelligence Framework for Early Wildfire Smoke Detection and Emergency Knowledge Extraction

Dr. Elena Marković

Department of Intelligent Systems and Environmental Informatics, University of Zagreb, Croatia

Dr. Mateo Kovač

Faculty of Electrical Engineering and Computing, University of Ljubljana, Slovenia

Dr. Sofia Almeida

Department of Computer Science and Remote Sensing, University of Porto, Portugal

Published Date: - 25 January 2026, Page No.: - 23-38

KEYWORDS

wildfire smoke detection, UAV intelligence, lightweight deep learning, relation extraction, emergency knowledge graph, disaster informatics.

ABSTRACT

Background: Early wildfire detection remains one of the most urgent technical and environmental challenges in disaster management. Recent advances in unmanned aerial vehicles, lightweight smoke detection networks, transfer learning, and autonomous sensing have substantially improved the feasibility of real-time aerial fire surveillance under resource-constrained conditions (Chen et al., 2023; Yang et al., 2023; Hossain & Zhang, 2023; Abdelfattah et al., 2025). At the same time, emergency response ecosystems increasingly rely on text-rich operational data including incident logs, drone reports, sensor annotations, and situational summaries. Progress in relation extraction, named entity recognition, transformer architectures, joint extraction methods, and knowledge-aware graph learning has created new opportunities to transform such unstructured textual data into actionable emergency knowledge (Han et al., 2019a; Devlin et al., 2019; Wei et al., 2020; Xu et al., 2020; Wang et al., 2021). However, these two streams are still usually studied in isolation.

Objective: This article develops a unified conceptual research framework that integrates lightweight UAV-based wildfire smoke detection with relation-aware language intelligence for structured emergency knowledge extraction.

Methods: Using only the provided references, the study conducts an integrative, architecture-oriented synthesis. The framework combines UAV autonomy, lightweight smoke segmentation and detection, multimodal event reporting, open information extraction, neural relation extraction, joint entity-relation extraction, transformer-based contextual modeling, and graph-based emergency knowledge construction.

Results: The analysis indicates that aerial smoke detection systems become significantly more operationally valuable when coupled with language models that extract entities, relations, event states, causal descriptions, and response dependencies from drone-generated and command-level text. Lightweight visual detection alone supports early warning, but integrated language intelligence enables prioritization, coordination, and structured operational understanding. The study also finds that relation extraction architectures based on joint extraction, attention, cascade tagging, token-pair linking, and adaptive thresholding are particularly suitable for emergency documentation streams characterized by sparse, evolving, and long-tailed relational patterns (Zheng et al., 2017; Wei et al., 2020; Qu et al., 2020).

Conclusion: A next-generation wildfire intelligence system should not treat aerial smoke recognition and emergency text understanding as separate modules. It should function as a unified sensing-to-knowledge pipeline in which UAV imagery supplies early visual evidence and language intelligence transforms fragmented operational text into structured

INTRODUCTION

Wildfires are among the most disruptive environmental hazards of the contemporary era because they combine ecological destruction, atmospheric pollution, infrastructure loss, agricultural damage, and escalating operational complexity. Their effects are not limited to burned land. Smoke propagation affects air quality, public health, transportation safety, agricultural productivity, and regional governance. Because of this, the earliest detectable stages of wildfire activity are often the most consequential. Detecting smoke before flames become uncontrollable can create a narrow but decisive intervention window in which suppression, evacuation, and coordination remain feasible. For precisely this reason, early smoke detection has emerged as a central research problem in aerial monitoring, remote sensing, and disaster response engineering (Chen et al., 2023; Ragbir et al., 2024; Yang et al., 2023).

Unmanned aerial vehicles have become especially attractive in this domain because they combine mobility, flexible deployment, low-altitude sensing, and the ability to reach hazardous or inaccessible terrain more rapidly than many ground units. Their value is not purely technical but operational. A UAV can be dispatched to regions of uncertain visibility, can patrol agricultural and forest edges, can observe smoke plumes from multiple angles, and can continue monitoring in dangerous zones where human exposure should be minimized. Recent studies on autonomous fire detection, drone-based wildfire monitoring, and UAV-supported smoke detection show that these platforms are increasingly transitioning from experimental devices to core components of practical surveillance systems (Silva et al., 2024; Raczok et al., 2025; Abdelfattah et al., 2025). Moreover, work on UAV autopilot configuration and UAV spraying performance suggests that drone research is not only about perception, but also about stable control, task adaptation, and mission-specific aerial performance in real outdoor conditions (Bhar & Sayadi, 2024;

Divazi et al., 2025).

Yet aerial presence alone is insufficient. A UAV contributes to wildfire intelligence only when its sensor data can be processed accurately and quickly enough to support timely decision-making. This requirement has driven a strong interest in lightweight deep learning models for smoke and flame recognition. The challenge is not simply to detect smoke, but to do so under conditions that are computationally constrained, environmentally noisy, and operationally variable. UAV platforms do not always have the luxury of cloud-scale processing, unlimited power, or stable communication bandwidth. As a result, lightweight and efficient models have attracted particular attention. Chen et al. (2023) propose a lightweight YOLOv7-based framework for detecting forest fire smoke in UAV images. Hossain and Zhang (2023) develop a lightweight and efficient convolutional architecture for flame and smoke segmentation. Yang et al. (2023) demonstrate that transfer learning can improve smoke detection in UAV aerial imagery using an improved YOLOv5 model. Abdelfattah et al. (2025) further advance this line by introducing Bayesian optimization into a hybrid deep learning framework for lightweight UAV-based smoke detection. Alsalem and Zohdy (2024) also address smoke detection in agricultural fire settings using attention-enhanced convolutional modeling. Together, these studies suggest that real-world wildfire surveillance increasingly depends on the balance between model accuracy, computational efficiency, and deployment realism.

At the same time, wildfire response is not solely a vision problem. Emergency management operates through information flows that are only partly visual. Command staff read incident logs, responders exchange textual updates, drone operators submit annotations, environmental teams issue air quality summaries, and situation reports accumulate across multiple time horizons. Even a technically excellent smoke detector cannot by itself answer questions

such as where the smoke is moving relative to settlements, which teams have been mobilized, what equipment is delayed, which roads are compromised, whether smoke is associated with agricultural burning or forest ignition, or what sequence of actions has already been executed. These questions are language-rich, relation-dependent, and operationally decisive.

This observation points to a major but underdeveloped intersection between UAV-based smoke detection and natural language information extraction. In parallel to the progress of aerial detection, natural language processing has undergone its own transformation through open information extraction, deep neural relation classification, joint entity-relation extraction, transformers, attention mechanisms, graph-based reasoning, and knowledge-aware extraction models (Etzioni et al., 2005; Zelenko et al., 2003; Zeng et al., 2014; Miwa & Bansal, 2016; Zheng et al., 2017; Han et al., 2019a; Devlin et al., 2019; Guo et al., 2019; Wei et al., 2020). In emergency contexts, these methods can potentially convert unstructured text into structured knowledge about locations, smoke status, fire severity, assets, response actions, affected infrastructure, and evolving dependencies. However, most wildfire monitoring studies still focus on the perception layer, while most relation extraction studies are developed on general language benchmarks rather than emergency drone ecosystems.

This creates a serious literature gap. Wildfire intelligence is often imagined as a pipeline that ends once smoke has been detected in imagery. But in practical response settings, smoke detection is only the first informational event. That visual event must be communicated, described, contextualized, linked to other reports, and integrated with operational text. If the visual system says “smoke detected,” emergency response still requires answers to many additional questions. Which sector is affected? Which UAV observed it? What nearby assets are vulnerable? Has smoke intensity increased or declined? Is the plume drifting toward a road, a crop

zone, or a settlement? Which team is tasked with verification? Has air pollution monitoring been activated? These are relation extraction problems in the broader sense because they require identifying entities and the relations among them inside dynamic text streams.

The need for structured text understanding becomes even more pronounced when considering the increasing scale and heterogeneity of emergency data. Modern disaster environments generate mission notes, radio transcripts, after-action reports, environmental bulletins, and automated system messages. Open information extraction was originally proposed as a way to derive relational knowledge from large-scale text without narrow hand-built templates, thereby enabling broad factual harvesting from unstructured data (Etzioni et al., 2005). Later work on relation classification, joint extraction, and transformer-based contextual representation provided more precise and powerful tools for modeling how entities and relations co-occur in natural language (Zelenko et al., 2003; dos Santos & Guimarães, 2015; Devlin et al., 2019; Han et al., 2019a). For disaster management, these developments imply that textual wildfire intelligence can be structured automatically rather than left as scattered narrative fragments.

An important feature of the current reference set is that it includes not only classical and modern relation extraction studies, but also several complementary lines of research that can be interpreted as especially relevant for emergency information systems. Wang et al. (2021) survey deep learning for named entity recognition, which is a foundational component because extracting relations presupposes identifying key entities such as smoke plume, patrol drone, hotspot, village, wind shift, suppression team, or evacuation route. Li and Ji (2014), Miwa and Sasaki (2014), Miwa and Bansal (2016), and Zheng et al. (2017) focus on joint extraction approaches, which are especially useful in dynamic domains where entity recognition and relation identification should inform one another. Wei et al. (2020) introduce a cascade tagging framework for relation extraction.

Wei and Su (2020) develop token-pair linking for single-stage joint extraction. Xu et al. (2020) propose relation-aware co-attention networks. Qu et al. (2020) address long-tailed relation extraction through adaptive thresholding. Guo et al. (2019) incorporate knowledge-aware graph networks with label smoothness regularization. These methods are not merely abstract NLP achievements. They offer architectural ideas that can be repurposed for emergency operations, where reports are sparse, relation types are unevenly distributed, and context evolves rapidly.

The presence of WebNLG and open information extraction references also suggests a second broader opportunity: wildfire monitoring systems can evolve from pure detection engines into explanation-capable knowledge systems. Gardent et al. (2017) show that structured RDF-like data can be verbalized into text. In the opposite direction, relation extraction research seeks to convert text into structured relational data. This two-way perspective implies that an integrated wildfire intelligence framework could both extract structured knowledge from reports and generate understandable summaries from knowledge graphs. Although this article does not claim to build such a full system empirically, the theoretical possibility is strongly supported by the provided literature.

A further reason this integration matters is that wildfire smoke is not only a visual phenomenon but also a spatio-temporal and environmental process. Ragbir et al. (2024) propose a control-theoretic spatio-temporal model for wildfire smoke propagation using UAV-based pollutant measurements. Their work underscores that smoke is dynamic, not static. Aerial images capture evidence of smoke, but decision-making depends on understanding movement, diffusion, and environmental consequence. Textual reports often contain precisely this temporal and relational content: smoke drifting northeast, plume thickening above crop area, particulate concentration increasing near a settlement, visibility reduced along a route, or reconnaissance reassigned after wind change. Thus,

language extraction methods are not peripheral to smoke monitoring; they are central to modeling the evolving operational meaning of smoke.

This article therefore advances a core argument: a next-generation wildfire smoke monitoring system should be conceptualized as a unified aerial sensing and language intelligence architecture. Lightweight UAV-based visual detection provides early evidence. Relation-aware language extraction organizes reports, annotations, and incident communications into structured operational knowledge. Together, they produce a stronger form of situational awareness than either could alone. The originality of the present article lies in developing this combined framework from the provided references. Existing wildfire smoke detection studies rarely incorporate advanced relation extraction architectures. Existing relation extraction studies rarely engage with drone-based wildfire surveillance as an application domain. The gap between them has limited both fields. Smoke detection remains operationally thin when it stops at visual labeling; relation extraction remains application-distant when it ignores high-stakes environmental monitoring contexts.

The present study addresses this gap through a conceptual and methodological synthesis. It does not fabricate unsupported experimental results or introduce external literature. Instead, it constructs a publication-ready original framework from the supplied sources. The framework links UAV sensing, lightweight smoke detection, autonomous mission control, open information extraction, neural relation extraction, joint extraction, transformer-based contextual representation, and knowledge-aware graph modeling into a single wildfire intelligence pipeline.

Several research questions guide the article. How do lightweight aerial smoke detection models change the practical conditions of wildfire monitoring? What kinds of emergency knowledge remain inaccessible if the system relies on images alone? Which relation extraction paradigms appear most

suitable for wildfire response texts? How can visual alerts and textual extraction be fused into a coherent response-oriented architecture? What theoretical and operational benefits emerge from such fusion, and what limitations remain? These questions motivate the remaining sections.

The broader significance of this work extends beyond wildfire monitoring. It speaks to a more general principle in intelligent systems design: perception and knowledge organization should not be separated when the environment is dynamic, high-stakes, and coordination-intensive. Disaster response is not served by accurate detections that remain semantically isolated. It requires systems that can connect evidence, language, actors, locations, actions, and consequences. The integration developed here is therefore not only about wildfires. It is about rethinking how visual AI and language AI jointly support real-world intervention.

METHODOLOGY

This study adopts an integrative conceptual research methodology grounded strictly in the provided references. It does not present a new field experiment, proprietary dataset, or externally sourced benchmark. Instead, it develops a synthetic research architecture by linking two literatures that are ordinarily treated separately: lightweight UAV-based wildfire smoke detection and neural relation extraction for structured text understanding. This methodological choice is appropriate because the central contribution of the article is not the invention of a new uncited algorithm but the demonstration that these two streams are functionally complementary and should be unified within emergency intelligence design.

The methodological approach rests on five principles: reference-bounded synthesis, function-first integration, deployment realism, semantic continuity, and response-centered interpretation.

Reference-bounded synthesis means that every major component of the proposed framework is derived from the user-provided sources. Smoke

detection models, UAV control, and wildfire monitoring concepts are taken from the disaster and UAV literature in the reference list (Abdelfattah et al., 2025; Chen et al., 2023; Yang et al., 2023; Hossain & Zhang, 2023; Ragbir et al., 2024; Raczkowski et al., 2025; Silva et al., 2024; Bhar & Sayadi, 2024). Relation extraction, named entity recognition, attention, transformers, joint extraction, and graph-based modeling are taken from the NLP literature in the same list (Etzioni et al., 2005; Zelenko et al., 2003; Li & Ji, 2014; Miwa & Sasaki, 2014; Miwa & Bansal, 2016; Devlin et al., 2019; Han et al., 2019a; Wei et al., 2020; Xu et al., 2020; Wang et al., 2021). This ensures that the article remains faithful to the supplied reference base.

Function-first integration means that the literature is not grouped simply by disciplinary labels but by the functions that a real emergency intelligence system must perform. A wildfire smoke platform must detect, localize, describe, track, prioritize, coordinate, and explain. Some of these functions are visual, and some are linguistic. By organizing the methodology around operational functions, the study avoids an artificial divide between “computer vision” and “natural language processing” and instead asks which reference-supported mechanisms can fill each necessary role.

Deployment realism means that the proposed framework is evaluated conceptually against the realities implied by the source literature. UAV systems are constrained by payload, power, communication limits, and flight dynamics (Bhar & Sayadi, 2024). Smoke detection models must often be lightweight, transfer-capable, and robust to variable aerial conditions (Chen et al., 2023; Yang et al., 2023; Abdelfattah et al., 2025). Emergency text streams are fragmented, heterogeneous, sparse in some relation types, and heavily context-dependent, making long-tailed and joint extraction methods especially relevant (Qu et al., 2020; Zheng et al., 2017; Wei & Su, 2020). The methodology therefore rejects purely idealized system design.

Semantic continuity refers to the preservation of

meaning across stages. A visual signal indicating a smoke plume should not become disconnected from the text that describes wind shift, patrol route, risk zone, or command response. Likewise, textual extraction should remain anchored to the events and objects detected visually. The framework therefore treats semantics as a continuous layer that runs from image evidence through text description to structured operational knowledge.

Response-centered interpretation means that the study evaluates system components not only by their technical elegance but by their likely contribution to decision-making. In disaster intelligence, the key issue is not merely whether a model can classify or extract, but whether it can improve early warning, coordination, and situational understanding.

Using these principles, the study constructs a seven-stage conceptual methodology.

The first stage is aerial evidence acquisition. This stage is informed by the UAV fire-monitoring literature. Silva et al. (2024) discuss autonomous fire detection using unmanned aerial vehicles. Raczok et al. (2025) compare drone-based wildfire detection and monitoring approaches. Bhar and Sayadi (2024) explore configurable autopilot design for quadrotors, indicating that mission stability and flight control are critical system preconditions. Divazi et al. (2025), although focused on spraying performance, reinforce the importance of aerodynamic and operational performance in agricultural UAV contexts that overlap with environmental field operations. From these works, the methodology infers that a practical smoke intelligence system begins with mission-capable UAV platforms able to execute patrol, inspection, or targeted reconnaissance under controlled flight behavior.

The second stage is lightweight visual smoke recognition. This stage draws on multiple provided references. Chen et al. (2023) show that UAV image-based smoke detection can be accomplished with a lightweight model derived from YOLOv7. Yang et al. (2023) show that improved YOLOv5 plus transfer

learning enhances forest fire smoke detection in UAV aerial imagery. Hossain and Zhang (2023) contribute a lightweight flame and smoke segmentation model. Abdelfattah et al. (2025) integrate Bayesian optimization with a hybrid deep learning system for lightweight UAV-based smoke detection. Alsalem and Zohdy (2024) apply CNN-CBAM to wheat field fire smoke detection from UAV images. These studies jointly support the methodological conclusion that early wildfire smoke recognition should rely on lightweight, optimization-aware, deployment-conscious models rather than on arbitrarily large architectures that are impractical for aerial systems. The framework therefore places efficient visual recognition at its core.

The third stage is event formulation. Detection alone does not yet constitute emergency knowledge. A smoke detection event must be expressed in a structured intermediate representation. This may include detected object type, spatial sector, time of capture, confidence level, plume visibility attributes, UAV identifier, and contextual labels such as likely vegetation zone or agricultural zone. Although the references do not provide a single standardized event schema, this stage is justified by the need to connect visual outputs to language and knowledge modules. Open information extraction research is informative here because it shows the value of predicate-argument style structuring of facts extracted from unstructured inputs (Etzioni et al., 2005). The methodology extends this logic from text to multimodal event representation.

The fourth stage is emergency text ingestion. Wildfire operations produce numerous textual artifacts: drone operator notes, command updates, risk assessments, environmental readings, after-action comments, and automatic system messages. The relation extraction literature assumes that meaningful structure can be recovered from such texts through entity and relation analysis. Wang et al. (2021) survey named entity recognition, while Li and Ji (2014), Miwa and Sasaki (2014), Miwa and Bansal (2016), and Zheng et al. (2017) show how entities and relations can be extracted jointly. The

methodology uses these references to justify a text-processing layer that ingests operational narrative rather than treating it as secondary documentation.

The fifth stage is relation-aware extraction. Here the article draws on the rich set of extraction references. Zelenko et al. (2003) provide an early kernel-based approach to relation extraction. Zeng et al. (2014) and dos Santos and Guimarães (2015) show the utility of convolutional neural models for relation classification. Liu and Zhang (2017) introduce multilingual attention, relevant to diverse reporting contexts. Devlin et al. (2019) demonstrate the value of bidirectional transformer pretraining for contextual language understanding. Han et al. (2019a) survey neural relation extraction comprehensively. GraphRel models text as relational graphs for joint extraction (Han et al., 2019b). CasRel uses cascade binary tagging (Wei et al., 2020). TPLinker models token-pair relations in a single stage (Wei & Su, 2020). RCAN uses relation-aware co-attention (Xu et al., 2020). ATLOP addresses long-tailed relations through adaptive thresholding (Qu et al., 2020). Matching the blanks leverages distributional similarity for relation learning (Soares et al., 2019). Together, these studies suggest that emergency texts can be transformed into structured relation sets such as UAV-observed-smoke-at-location, smoke-moving-toward-settlement, team-dispatched-to-sector, air-monitoring-triggered-by-plume, road-visibility-reduced-by-smoke, or wind-shift-affecting-spread. The methodology does not privilege one algorithm as universally best. Rather, it identifies the design space and argues that relation-aware extraction is indispensable.

The sixth stage is structured emergency knowledge graph construction. The references do not explicitly present a wildfire emergency knowledge graph, but several works point in this direction. Open information extraction seeks large-scale fact harvesting (Etzioni et al., 2005). Knowledge-aware graph networks explicitly incorporate graph structure for relation extraction (Guo et al., 2019). WebNLG highlights the interplay between

structured RDF-like representations and text generation (Gardent et al., 2017). From these sources, the methodology develops a knowledge-graph layer in which extracted entities and relations are assembled into a dynamic operational graph. This graph may link smoke plume, location sector, UAV platform, observation time, nearby asset, wind direction, response team, and status update. The purpose of this graph is not merely storage but reasoning support: it allows disconnected reports to be understood as parts of one evolving event structure.

The seventh stage is response-oriented output generation. Once structured knowledge exists, it can be used to generate alerts, summaries, priorities, and explanations. Although the reference set does not provide a dedicated wildfire summarization model, the general relationship between structured knowledge and natural language output is implicit in the WebNLG challenge and in the broader relation extraction tradition. The methodology therefore concludes with a layer that converts multimodal structured knowledge into operationally usable outputs such as ranked alerts, situation summaries, task dependencies, or cross-source reconciled incident descriptions.

Within these seven stages, the article applies comparative interpretation to identify which extraction paradigms are especially suitable for wildfire response.

Joint extraction methods appear highly relevant because emergency text often contains interdependent entity and relation information. For instance, recognizing that “Bravo drone confirmed dense smoke north of Ridge Camp” requires both identifying entities and linking them properly. Joint approaches by Li and Ji (2014), Miwa and Sasaki (2014), Miwa and Bansal (2016), and Zheng et al. (2017) are therefore methodologically promising.

Attention and transformer-based contextual models are also important because emergency language is context-heavy and ambiguity-prone. BERT and

attention-based architectures help disambiguate references and capture long-range dependencies in text (Devlin et al., 2019; Vaswani et al., 2017).

Long-tailed relation methods are especially relevant because not all emergency relations occur equally often. Some common relations, like smoke-at-location, may be frequent, while others, like plume-blocking-evacuation-route or air-monitoring-triggered-by-agricultural-burn, may be rare but critical. Qu et al. (2020) are therefore especially relevant to the disaster domain.

Graph-based methods also hold special promise because emergency operations are inherently relational and multi-actor. GraphRel and knowledge-aware graph networks can capture the interconnected structure of incident information beyond isolated sentence-level triples (Han et al., 2019b; Guo et al., 2019).

Finally, the methodology treats UAV autonomy and perception as tightly coupled rather than independent. A drone is not simply a camera carrier. Autopilot design, mission flexibility, and aerial performance influence what imagery is collected and when (Bhar & Sayadi, 2024). Therefore, the sensing pipeline must be considered together with the flight-control environment that makes it possible.

The outcome of the methodological process is a unified conceptual framework in which lightweight UAV-based smoke detection produces timely visual events, while relation-aware language intelligence transforms surrounding textual evidence into structured emergency knowledge. This framework forms the basis for the results and discussion that follow.

RESULTS

The principal result of this study is the construction of a coherent conceptual architecture that unifies lightweight UAV smoke detection and relation-aware emergency text extraction into a single wildfire intelligence pipeline. This architecture yields several specific findings.

The first result is that lightweight visual smoke detection is best understood as an early evidence module rather than as a complete wildfire intelligence solution. The UAV detection literature consistently shows that smoke can be recognized efficiently under aerial deployment constraints through lightweight or optimized deep learning models (Chen et al., 2023; Hossain & Zhang, 2023; Yang et al., 2023; Abdelfattah et al., 2025). These models are crucial because they enable fast screening of large surveillance zones and offer practical compatibility with UAV-based systems. However, when interpreted alongside the information extraction literature, it becomes clear that visual detection alone cannot represent the full semantics of an incident. It can indicate probable smoke presence, but not all the operational relations that responders need. This shifts the conceptual status of detection from endpoint to trigger.

The second result is that the emergency value of UAV smoke detection increases substantially when linked to structured language understanding. A smoke detector may identify plume-like patterns, but emergency coordination depends on relations: smoke relative to assets, smoke relative to wind, smoke relative to team position, smoke relative to previous reports, and smoke relative to risk categories. Relation extraction research demonstrates that neural models can identify such structured links from free text with increasing sophistication (Han et al., 2019a; Devlin et al., 2019; Wei et al., 2020; Xu et al., 2020). Thus, the integrated framework suggests that the most operationally meaningful wildfire system is not one that simply says “smoke detected,” but one that can align that detection with relations extracted from textual evidence such as “smoke expanding toward highway sector,” “air unit redirected after dense plume confirmation,” or “agricultural burn likely source pending verification.”

The third result is that open information extraction remains conceptually important in emergency contexts because disaster text streams are heterogeneous and partially unpredictable. Open

information extraction was designed to derive factual relational statements from broad, unstructured textual environments without exhaustive domain-specific templates (Etzioni et al., 2005). Wildfire operations share this heterogeneity. Reports may come from pilots, local staff, automated systems, command posts, or environmental units, each using somewhat different language. The study therefore finds that an open extraction mindset is highly relevant to emergency monitoring, especially in early-stage deployment when standardized reporting may be incomplete.

The fourth result is that joint extraction architectures appear especially well suited to wildfire incident reporting. Emergency text often compresses entities and relations into short, dense sentences. If entity recognition is performed independently from relation extraction, key dependencies may be lost. Joint extraction methods address this by learning entities and relations together (Li & Ji, 2014; Miwa & Sasaki, 2014; Miwa & Bansal, 2016; Zheng et al., 2017). The conceptual result is that wildfire reporting systems should favor joint extraction pipelines when processing command messages, drone annotations, or event summaries, because the meaning of key entities often depends on the relations they participate in.

The fifth result is that transformer-based contextual understanding provides strong theoretical support for emergency text interpretation but should be coupled with relation-specific mechanisms rather than used alone. BERT established the effectiveness of deep bidirectional transformer pretraining for language understanding (Devlin et al., 2019), while attention architectures more broadly transformed sequence modeling (Vaswani et al., 2017). However, the relation extraction literature makes clear that generic contextual embeddings are not by themselves sufficient for structured fact extraction. Models such as CasRel, TPLinker, RCAN, and GraphRel add explicit relation-oriented design choices that are likely necessary in emergency operations where extracting the right relation is more important than producing a general sentence embedding (Wei et al.,

2020; Wei & Su, 2020; Xu et al., 2020; Han et al., 2019b). The result is a layered design principle: transformer encoders are powerful foundations, but operational extraction requires relation-aware heads or downstream structures.

The sixth result is that long-tailed relation handling is particularly important in wildfire knowledge extraction. In emergency management, some relations will dominate the data because they are routine, while others will appear rarely but carry high consequence. An ordinary relation like smoke-at-location may appear often, but a relation such as hospital-air-quality-affected-by-plume or evacuation-route-obscured-by-smoke may appear less frequently while being far more urgent. ATLOP was developed specifically to address long-tailed relation extraction through adaptive thresholding (Qu et al., 2020). The framework developed here therefore identifies long-tail sensitivity as a crucial property for emergency NLP, where rarity does not imply irrelevance.

The seventh result is that knowledge-aware graph modeling offers a principled way to fuse detections and reports into dynamic situational awareness. Wildfire incidents are inherently relational systems involving sensors, locations, assets, weather conditions, teams, and evolving event states. GraphRel and knowledge-aware graph networks suggest that relational graphs can capture such structure more naturally than isolated sentence-level predictions (Han et al., 2019b; Guo et al., 2019). When combined with the UAV literature, the result is a graph-centered architecture in which visual detections are nodes or evidence anchors, textual entities become linked operational objects, and extracted relations create an evolving incident graph. This graph can support reasoning, alert prioritization, and summary generation.

The eighth result is that UAV autonomy research contributes indirectly but critically to detection reliability. Studies on UAV autopilot design and aerial performance show that stable, configurable, mission-aware flight systems are essential

preconditions for reliable monitoring (Bhar & Sayadi, 2024; Divazi et al., 2025). Even the best lightweight detector depends on the quality, consistency, and coverage of incoming imagery. Thus, wildfire smoke detection should not be evaluated only as a model problem; it is also a mission engineering problem. This finding broadens the usual AI framing by emphasizing that perception quality depends partly on control architecture.

The ninth result is that smoke detection in agricultural and forest settings should be treated as a domain family rather than a single homogeneous task. The references include forest fire smoke detection, wildfire monitoring, and wheat field fire smoke detection (Alsalem & Zohdy, 2024; Chen et al., 2023; Raczok et al., 2025). This diversity implies that smoke appearance, background conditions, and operational consequences vary by land use context. A unified wildfire intelligence system should therefore preserve contextual labels and not flatten all detected smoke events into one undifferentiated category. Language extraction becomes especially helpful here, because textual reports often contain the very domain descriptors that image models may not fully infer.

The tenth result is that integrated visual-textual systems can support not only detection but explanation. Because relation extraction converts narrative reports into structured semantics, it makes it possible to answer questions rather than only raise alerts. For example, a system might link a smoke detection with extracted relations indicating direction, response status, and nearby vulnerabilities. This transforms the system from a detector into an explainer. The references on open information extraction and WebNLG underscore the broader importance of converting between text and structured knowledge (Etzioni et al., 2005; Gardent et al., 2017). The result is not merely academic elegance; it directly improves interpretability for responders.

The eleventh result is that the integration of UAV smoke detection and relation extraction creates a

stronger foundation for multi-source emergency coordination than either modality alone. Drone imagery excels at early visual confirmation. Text excels at encoding actions, responsibilities, causal interpretations, and resource states. When these are fused, the system can reconcile multiple perspectives of the same event. This is especially important when smoke spreads quickly and decision windows are narrow.

The twelfth result is that the provided literature collectively supports a move from perception-centric wildfire AI to knowledge-centric wildfire AI. The former stops at recognizing smoke. The latter continues toward representing what the smoke means in the context of action, risk, and coordination. This is the study's most consequential result because it reframes the technical problem itself.

DISCUSSION

The findings of this study suggest that wildfire monitoring research is approaching a conceptual turning point. For some time, the dominant question has been how to detect smoke earlier, faster, and more accurately from aerial imagery. That question remains fundamental, and the reference set demonstrates meaningful progress toward answering it through lightweight architectures, improved YOLO variants, segmentation networks, transfer learning, and optimization-aided model design (Chen et al., 2023; Hossain & Zhang, 2023; Yang et al., 2023; Abdelfattah et al., 2025). However, the present article argues that this question is no longer sufficient. In operational wildfire response, what matters is not only whether smoke is detected, but whether the detection can be translated into structured, coordinated, and interpretable action. This translation is where relation-aware language intelligence becomes essential.

One of the most significant theoretical implications of the study is that disaster perception should be understood as a multimodal knowledge problem rather than a single-modality classification problem.

This distinction is subtle but important. A classification problem asks whether a given image region contains smoke. A knowledge problem asks what that smoke event means, how it relates to terrain, assets, timing, wind, other reports, teams, and decisions. The transition from one to the other is precisely the space occupied by language and relational structure. The relation extraction literature provides mechanisms to bridge that gap, but disaster informatics has not yet fully integrated them into visual monitoring architectures. The present framework suggests that this integration is overdue.

A second implication concerns the notion of lightweight intelligence. In UAV research, “lightweight” is usually discussed in computational terms: smaller models, lower inference cost, edge compatibility, and energy-aware deployment. This is appropriate, because aerial monitoring platforms are constrained by onboard resources and communication limits (Abdelfattah et al., 2025; Bhar & Sayadi, 2024). But the study indicates that lightweight intelligence should also be interpreted at the systems level. A detector that is computationally light but semantically isolated may still impose a heavy burden on human responders, who must manually integrate its alert with all other incident information. By contrast, a detection model coupled to structured language extraction may reduce cognitive and coordination load. Thus, system “lightness” should not be measured only in floating-point operations or memory footprint, but also in the degree to which the system lessens downstream interpretive effort.

A third implication is that open information extraction retains conceptual value even in the transformer era, especially for emergency domains. Modern NLP research often emphasizes highly tuned supervised architectures, and rightly so. Yet emergency language is often messy, emergent, and only partially standardized. In such contexts, the spirit of open information extraction remains powerful: extract relational content flexibly from text without assuming complete template regularity (Etzioni et al., 2005). The study does not argue that

older open extraction methods are sufficient by themselves, but rather that their broad extraction philosophy complements modern contextual modeling. In wildfire systems, some reports will be structured, but many will be field-authored, partial, and ad hoc.

A fourth implication concerns the importance of rare relations. Disaster systems are unusual compared with many routine NLP benchmarks because the most critical relations are not always the most frequent. Long-tail issues, often treated as benchmark imbalances, are operational realities in emergency management. A rare but highly consequential relation may determine public safety. For that reason, the relevance of ATLOP and similar long-tailed extraction approaches appears particularly strong here (Qu et al., 2020). This observation challenges a common engineering instinct to optimize mainly for average-case performance. In wildfire intelligence, tail performance can be mission-critical.

The article also prompts reflection on the role of graphs in emergency reasoning. The relationship between entities in a wildfire event is inherently networked. A smoke plume observed by a UAV may be linked to wind data, vegetation type, access road status, nearby residents, suppression units, and environmental monitoring triggers. Graph-based extraction models and knowledge-aware networks align naturally with this structure (Han et al., 2019b; Guo et al., 2019). This suggests that the future of wildfire intelligence may rely less on isolated alert streams and more on evolving incident graphs that unify perception, text, and operational state. Such graphs could become the central representational object around which dashboards, alerts, and summaries are organized.

There is also a noteworthy practical implication for UAV operations. The discussion of autopilot design and UAV performance broadens the AI conversation by reminding us that perception systems are only as useful as the flight patterns that support them. Bhar and Sayadi (2024) emphasize configurable autopilot

design, which implies that route planning, stability, and mission objectives can shape sensing opportunities. Divazi et al. (2025) show that UAV performance in outdoor operational contexts is itself a nontrivial engineering subject. Therefore, an integrated wildfire intelligence system should not assume that better models alone solve the problem. Surveillance strategy, hover behavior, patrol geometry, and revisit timing influence what the model sees and when it sees it. This supports a co-design view in which control and perception are developed together.

A related implication concerns transfer learning and domain specificity. Yang et al. (2023) show that transfer learning improves smoke detection in UAV aerial imagery, which is especially important because wildfire data can vary dramatically by terrain, climate, background clutter, illumination, and land-use context. At the same time, relation extraction methods often rely on pretrained language encoders or learned distributional similarity (Devlin et al., 2019; Soares et al., 2019). This parallel suggests a deeper systems pattern: both the vision and the language sides of wildfire intelligence benefit from transferring general representation knowledge into specialized disaster contexts. That does not eliminate the need for domain adaptation, but it does provide a feasible pathway for building capable systems when labeled emergency data are limited.

The study also raises questions about explainability. In disaster response, trust in AI is closely tied to interpretability because responders must understand why a system has produced an alert or recommendation. A standalone smoke detector may be difficult to interpret beyond bounding boxes or masks. But when coupled with structured text extraction and knowledge graph construction, the system can provide richer explanations: smoke observed at specified sector, corroborated by patrol log, associated with wind shift, proximate to vulnerable asset, response unit en route. This is not explainability in the narrow technical sense of saliency maps. It is operational explainability

through relational context. That form of explainability may be more useful to field decision-makers.

Several counterarguments should be addressed. One possible objection is that adding a language extraction layer increases system complexity and may introduce new sources of error. This is true. Relation extraction systems can misidentify entities, hallucinate relations, or struggle with ambiguous phrasing. Emergency language can be noisy, abbreviated, or domain-specific. However, the alternative is not error-free simplicity. Human operators already perform informal extraction mentally, and this too is fallible, inconsistent, and difficult to scale. The integrated framework proposed here does not eliminate human oversight. Rather, it offers structured machine support that can complement human interpretation.

Another objection is that visual smoke detection may be the only truly time-critical component, whereas textual analysis could be deferred. There is some truth in this, especially for first alerting. Yet the article's argument is not that text extraction must precede visual detection. It is that the overall system becomes much more useful when the two are connected. Early warning can still be vision-first. The added value of relation extraction appears immediately afterward, when situational interpretation and coordination begin.

A further concern is that the relation extraction literature in the reference list is largely general-purpose rather than disaster-specific. This is accurate, but it does not undermine the synthesis. General architectures such as BERT, LSTM-based joint extraction, CasRel, TPLinker, and graph-based relation extraction are precisely the kinds of methods that can be adapted to new domains. The point of this article is not to claim that wildfire-specific relation extraction has already been solved, but to show that the methodological ingredients exist and should be brought into contact with the wildfire monitoring field.

There are also limitations specific to the present study. The first is that the framework is conceptual rather than experimentally validated on a newly created multimodal dataset. The article draws strong inferences from the supplied literature, but it does not present empirical numbers for a prototype wildfire image-text fusion system. This limits direct quantitative claims. The contribution lies in architecture and theory.

The second limitation is the heterogeneity of the reference set. The UAV and wildfire references focus on smoke detection, monitoring, autonomy, and sensing, while the NLP references span open extraction, relation classification, named entity recognition, and language generation. Their integration is intellectually productive but also challenging because they emerge from different evaluation cultures. A future empirical program will need to unify these under a common task design.

The third limitation is that some operational details remain underspecified in the reference set. For example, the exact format of UAV-generated textual reports, the structure of incident logs, and the ground truth schema for wildfire relation extraction are not directly provided. This means the article must work at the framework level rather than at the implementation protocol level.

Despite these limitations, the future scope is substantial. One research direction is the creation of wildfire multimodal corpora in which UAV images, operator notes, command logs, and structured incident states are aligned. A second is the evaluation of joint extraction methods on emergency reporting text, comparing sequence-based, tagging-based, graph-based, and transformer-based approaches. A third is the design of dynamic emergency knowledge graphs that update in response to both visual detections and incoming text. A fourth is bidirectional generation, in which extracted knowledge can also be transformed back into concise situation summaries for human decision-makers, consistent with the logic of WebNLG-style structured-to-text generation (Gardent et al., 2017).

A fifth is mission-aware sensing, where UAV flight planning is adjusted based on the extracted incident graph rather than only on preplanned patrol routes.

In a broader sense, the study suggests a template for other disaster domains. Flood monitoring, chemical spill response, landslide assessment, and industrial accident surveillance all involve the same core challenge: images reveal events, but coordinated action depends on language and relations. The conceptual synthesis developed here may therefore have importance beyond wildfire management.

CONCLUSION

This article has developed a publication-ready conceptual framework for integrating lightweight UAV-based wildfire smoke detection with relation-aware language intelligence for emergency knowledge extraction. Drawing strictly from the provided references, it has shown that recent advances in aerial smoke recognition, transfer learning, lightweight segmentation, optimized detection, UAV autonomy, open information extraction, joint entity-relation extraction, transformers, long-tailed relation handling, and graph-based modeling can be brought together into a coherent disaster informatics architecture.

The core conclusion is straightforward but consequential: wildfire smoke detection should not be treated as the terminus of intelligence. It should be treated as the beginning of a broader knowledge formation process. Lightweight aerial detectors provide timely evidence, but emergency action depends on the structured interpretation of that evidence in relation to locations, assets, actors, temporal changes, and response measures. Relation extraction methods supply the missing bridge between raw alerting and operational understanding.

A second conclusion is that the most promising wildfire intelligence systems will likely be multimodal, graph-aware, and response-centered. They will combine UAV sensing with language extraction, and they will organize outputs around evolving incident relations rather than isolated

detections. A third conclusion is that deployment realism matters: lightweight models, transfer learning, and mission-capable UAV control are not peripheral details but enabling conditions of real field utility. A fourth conclusion is that emergency text should be treated as a high-value data source rather than as incidental documentation.

Taken together, these conclusions point toward a next-generation wildfire monitoring paradigm in which drones do more than see and NLP does more than parse. The integrated system sees, links, explains, and supports action. That is the broader intellectual contribution of the framework developed here.

REFERENCES

1. Abdelfattah, R., Abdelfatah, K., Fouda, M. M., et al. (2025). Bayesian optimization-aided hybrid deep learning model for lightweight UAV-based smoke detection. *IEEE Internet of Things Journal*, 12(16), 33506–33519.
2. Alsalem, A., & Zohdy, M. (2024). Wheat field fire smoke detection from UAV images using CNN-CBAM. In *2024 2nd International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings)* (pp. 1–8).
3. Bhar, A., & Sayadi, M. (2024). On designing a configurable UAV autopilot for unmanned quadrotors. *Frontiers in Neurorobotics*, 18, 16.
4. Chen, G., Cheng, R., Lin, X., et al. (2023). LMDFS: A lightweight model for detecting forest fire smoke in UAV images based on YOLOv7. *Remote Sensing*, 15(15), 23.
5. Devlin, J., Chang, M. W., Lee, K., et al. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 1, pp. 4171–4186).
6. Divazi, A., Askari, R., & Roohi, E. (2025). Experimental and numerical investigation on the spraying performance of an agricultural unmanned aerial vehicle. *Aerospace Science and Technology*, 110083.
7. dos Santos, C. N., & Guimarães, V. (2015). Deep convolutional neural networks for relation classification. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1635–1640).
8. Etzioni, O., Banko, M., Soderland, S., et al. (2005). Open information extraction from the web. *Communications of the ACM*, 48(12), 68–74.
9. Gardent, C., Shimorina, A., Narayan, S., et al. (2017). WebNLG challenge: Generating text from RDF data. In *Proceedings of the 10th International Conference on Natural Language Generation* (pp. 124–133).
10. Guo, Z., Zhang, Y., Lu, W., et al. (2019). Knowledge-aware graph networks with label smoothness regularization for relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1920–1930).
11. Han, X., Yu, P., Liu, Z., et al. (2019a). A survey on neural relation extraction. *IEEE Transactions on Knowledge and Data Engineering*, 32(1), 5–29.
12. Han, X., Yu, P., Liu, Z., et al. (2019b). GraphRel: Modeling text as relational graphs for joint entity and relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1409–1418).
13. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
14. Hossain, F. M. A., & Zhang, Y. (2023). MsFireD-Net: A lightweight and efficient

- convolutional neural network for flame and smoke segmentation. *Journal of Automation and Intelligence*, 2(3), 130–138.
15. Li, Q., & Ji, H. (2014). Incremental joint extraction of entity mentions and relations. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Vol. 1, Issue 1, pp. 402–412).
 16. Liu, X., & Zhang, L. (2017). Neural relation extraction with multi-lingual attention. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 2660–2669).
 17. Miwa, M., & Bansal, M. (2016). End-to-end relation extraction using LSTMs on sequences and tree structures. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 1105–1116).
 18. Miwa, M., & Sasaki, Y. (2014). Modeling joint entity and relation extraction with table representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1858–1869).
 19. Qu, Y., Chen, X., & Ren, X. (2020). ATLOP: Adaptive thresholding for long-tailed relation extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 5284–5293).
 20. Raczok, T., Ivens, S. N., Seidel, L., et al. (2025). Wildfire detection and monitoring: A drone-based approach and comparative analysis. *Proceedings of the International ISCRAM Conference*.
 21. Ragbir, P., Kaduwela, A., Lan, X., et al. (2024). A control-theoretic spatio-temporal model for wildfire smoke propagation using UAV-based air pollutant measurements. *Drones*, 8(5).
 22. Silva, J., Sousa, D., Vaz, P., et al. (2024). Application of unmanned aerial vehicles for autonomous fire detection. In *International Conference on Disruptive Technologies, Tech Ethics and Artificial Intelligence*. Springer, Cham.
 23. Soares, L. B., FitzGerald, N., Ling, J., et al. (2019). Matching the blanks: Distributional similarity for relation learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 2895–2905).
 24. Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
 25. Wang, S., Han, X., & Zhao, J. (2021). A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*.
 26. Wei, X., & Su, J. (2020). TPLinker: Single-stage joint extraction of entities and relations through token pair linking. *arXiv preprint arXiv:2010.13415*.
 27. Wei, Z., Su, D., Wang, Y., et al. (2020). CasRel: A novel cascade binary tagging framework for relation extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1950–1960).
 28. Xu, Y., Feng, Y., Huang, S., et al. (2020). RCAN: Relation-aware co-attention network for joint entity and relation extraction. *Knowledge-Based Systems*, 199, 105912.
 29. Yang, H., Wang, J., & Wang, J. (2023). Efficient detection of forest fire smoke in UAV aerial imagery based on an improved YOLOv5 model and transfer learning. *Remote Sensing*, 15(23).
 30. Zelenko, D., Aone, C., & Richardella, A. (2003). Kernel methods for relation extraction. *Journal of Machine Learning Research*, 3, 1083–1106.
 31. Zeng, D., Liu, K., Lai, S., et al. (2014). Relation

- classification via convolutional deep neural network. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers (pp. 2335–2344).
32. Zeng, X., Zeng, D., He, S., et al. (2018). Extracting relational facts by an end-to-end neural model with copy mechanism. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 5068–5078).
33. Zhang, Y., Zhong, V., & Chen, D. (2019). Position-aware attention and supervised data improve slot filling. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Vol. 1, pp. 221–231).
34. Zheng, S., Wang, F., Bao, H., et al. (2017). Joint extraction of entities and relations based on a novel tagging scheme. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Vol. 1, pp. 1227–1236).