Machine learning in humanitarian logistics and supply chain management: application domains and quality assurance

Son Nguyen and *Nichapa Phraknoi*
The Business School, RMIT University, Ho Chi Minh City, Vietnam
Juliana Sutanto
Department of Human-Centred Computing, Faculty of Information Technology, Monash University, Melbourne, Australia, and
Irfan UlHaq and Hung Nguyen
The Business School, RMIT University, Ho Chi Minh City, Vietnam

Abstract

Purpose – Despite the growing integration of machine learning (ML) in humanitarian supply chain (HSC), herein HSCML, no prior review has established a systematic application domain taxonomy or examined model quality assurance, limiting sustainable adoption. This study aims to address three gaps: the absence of a systematic taxonomy for HSCML application domains, the lack of frameworks for HSCML data and model quality and the disconnect between technical development and operational deployability.

Design/methodology/approach – A sequential mixed-methods systematic literature review was used. In Phase 1, latent Dirichlet allocation topic modelling was applied to 302 HSCML journal papers to identify thematic clusters and derive a context–intervention–mechanism–outcome (CIMO) classification system. In Phase 2, a CIMO-guided systematic review analysed 93 HSCML model development studies across contextual, methodological and quality dimensions.

Findings – ML contributed to HSC through six domains: disaster consequence prediction, relief operations optimisation, real-time crisis intelligence, infrastructure monitoring and recovery, population mobility and evacuation and strategic HSC network planning. Research efforts are unevenly distributed across disaster types, management phases and regions, with fairness, explainability and uncertainty quantification systematically neglected across all domains. In response, three frameworks are established covering data quality dimensions, model quality dimensions and factors influencing model quality.

Originality/value – This is the first systematic examination of data and model quality in HSCML. It delivers a deployability-focused quality assurance framework and a data-centric research agenda addressing 14 knowledge gaps, providing practical references for rigorous HSCML development in line with humanitarian obligations.

Keywords Humanitarian supply chains, Artificial intelligence, Machine learning, Mixed-methods, Literature review

Paper type Literature review

1. Introduction

Humanitarian supply chain (HSC) focuses on the preparation for, response to and recovery from humanitarian crises, to save lives and alleviate the suffering of affected populations (Kembro *et al.*, 2024). The timely and effective assistance of HSC has been demonstrated in the preparation for and response to many disasters, benefiting millions of beneficiaries. Recovery and mitigation efforts, such as event and consequence forecasting (Kondraganti *et al.*, 2022), preparedness (Kumar *et al.*, 2022) and post-disaster recovery responses (Yağcıoğlu, 2025) were also a part of humanitarian operations. HSCs are guided primarily by non-profit objectives and must function under extreme uncertainty, resource constraints and time-critical conditions (de Camargo Fiorini *et al.*, 2022; Zhao *et al.*, 2025), requiring collaborative decision-making among stakeholders under different mandates and with diverse logistics capabilities (Altay *et al.*, 2023; Sentia *et al.*, 2023).

In such managerial and operational environments, artificial intelligence (AI) has emerged as a technology of significant potential, capable of delivering insights and recommendations tailored to operational situations. Among AI approaches,

The current issue and full text archive of this journal is available on Emerald Insight at: **https://www.emerald.com/insight/2042-6747.htm**

Journal of Humanitarian Logistics and Supply Chain Management
Emerald Publishing Limited [ISSN 2042-6747]
[DOI 10.1108/JHLSCM-02-2026-0045]

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

Received 4 March 2026
Revised 7 June 2026
Accepted 27 July 2026

machine learning (ML), in which models are trained on historical data to make predictions, classifications or optimisation decisions, can handle complex relationships to generate accurate outputs. As ML enables complex and high-value predictive and prescriptive analytics for decision-making, and given that the LSCM literature is overwhelmingly ML-based (Vlachos and Reddy, 2025), this review focuses specifically on ML in HSC (HSCML).

HSCML has evolved from early applications in disaster and risk management, driven primarily by computer science communities, towards a rapidly expanding and multi-disciplinary field encompassing operational logistics, real-time intelligence and network resilience design (e.g. damage assessment (Chen and Cho, 2022), situational awareness (Ullah *et al.*, 2021; Zhang *et al.*, 2024b), UAV for disaster reconnaissance and communication restoration (Zhu and Wang, 2023; Yağcıoğlu, 2025). Despite its emergence and advocates for adoption, the HSCML's foundational taxonomy and quality assurance have not been established. As evidenced in Table 1, existing HSC reviews address operational, organisational and sustainability dimensions without ML as a primary focus, and Vlachos and Reddy (2025) covered general ML-focused LSCM rather than humanitarian contexts. Where ML appears in humanitarian reviews, it is either incidental or prescribed as a future research direction, with no prior work establishing an application domain taxonomy or examining data and model quality dimensions.

Meanwhile, HSC stakeholders are struggling with adoption, with major concerns centred on data and model quality. Organisations like the UN Refugee Agency, the Office for the Coordination of Humanitarian Affairs (OCHA) and the International Committee of the Red Cross raised concerns about AI-based solutions producing problematic outputs stemming from data quality issues and algorithmic bias (OCHA, 2024). Empirical evidence confirms these concerns, documenting how opaque AI systems cause distrust and operational harm (Behl *et al.*, 2026), while an insufficient understanding of model quality dimensions and their trade-offs leaves operators ill-equipped to govern and maintain AI accountability in line with humanitarian principles and legal obligations (Dorsey, 2026). Addressing these challenges requires establishing quality assurance frameworks and a structured awareness of the factors that influence data and model quality. Through evidence-based insights into HSCML contexts, applications and outcomes, this study responds to these gaps through three key research questions (RQs):

RQ1. How was HSCML proposed and implemented across different contexts and application domains?

RQ2. How have data quality dimensions been recognised and attended to across HSCML application domains?

RQ3. What dimensions and factors characterise HSCML model quality?

RQ2 and *RQ3* target data and model quality as the mechanisms through which the gap between HSCML technical development and operational deployability manifests, directly identifying dimensions and factors whose systematic oversight obstructs adoption. Using latent Dirichlet allocation (LDA) topic modelling on 302 ML-relevant HSC papers and a context–intervention–mechanism–outcome (CIMO)-structured SLR of 93 HSCML model development studies, this review characterises HSCML across *six application domains*, detailing ML applications, implementations and HSC contributions in each context. This review systematically examines how *data quality* and *model quality* dimensions have been addressed and overlooked across domains. In response, three frameworks are established for *data quality dimensions*, *model quality dimensions* and *factors influencing model quality*, along with a deployability-focused, data-centric research agenda that addresses 14 identified knowledge gaps. Together, these contribute to principled, deployable HSCML development that aligns technical rigour with humanitarian obligations.

Table 1 Comparison of the existing relevant literature reviews and this study

Study	Scope	Methods	HSCML relevance	ML quality focus
Anjomshoae *et al*. (2022)	HSC performance measurement	Bibliometric + SLR	None	None
de Camargo Fiorini *et al*. (2022)	HSC human resource management	SLR	None	None
Maric *et al*. (2022)	Digital technologies in HSC	SLR	Limited: 2/110 papers addressed AI	None
Kumar *et al*. (2022)	HSC during COVID-19	Bibliometric + interviews	Limited: ML identified as a future direction only	None
Kondraganti *et al*. (2022)	Big data analytics in humanitarian operations	SLR	Limited: big data types and disaster categories	Partial: social media data concerns only
Sentia *et al*. (2023)	Logistics distribution in HSC	Thematic literature review	None	None
Altay *et al*. (2023)	Innovation in HSC	SLR	Limited: ML as one innovation example	None
Nawazish *et al*. (2023)	HSC sustainability	Bibliometric + SLR (TCCM)	Limited: ML as future research theme	None
Vlachos and Reddy (2025)	ML in general SCM	Bibliometric + SLR	Indirect: general SCM themes; not HSC-specific	None
This review	ML in HSC (HSCML)	LDA topic modelling + SLR	HSCML's 6 application domains are characterised	Data quality dimensions, model quality dimensions and influencing factors

Source(s): Authors' own work

The rest of the paper is organised as follows. Section 2 details the sequential mixed-methods SLR, including LDA topic modelling results that support the CIMO structure. Section 3 presents the results of the systematic literature review. Section 4 discusses key findings, research gaps and future directions, before the conclusion in Section 5.

2. Review methodology

This study uses a sequential mixed-methods SLR design, integrating quantitative and qualitative strands in sequence (Grant and Booth, 2009; Harden and Thomas, 2010). As no established HSCML application domain taxonomy existed before this study, Phase 1 deploys LDA topic modelling as the quantitative-dominant strand to identify the thematic structure of the HSCML corpus, particularly the application domains and expected outcomes of HSCML, from which an evidence-grounded CIMO classification framework is derived to structure Phase 2's CIMO-guided systematic review of model development studies, curated through uniformly applied quality and relevance criteria.

2.1 Data identification and collection

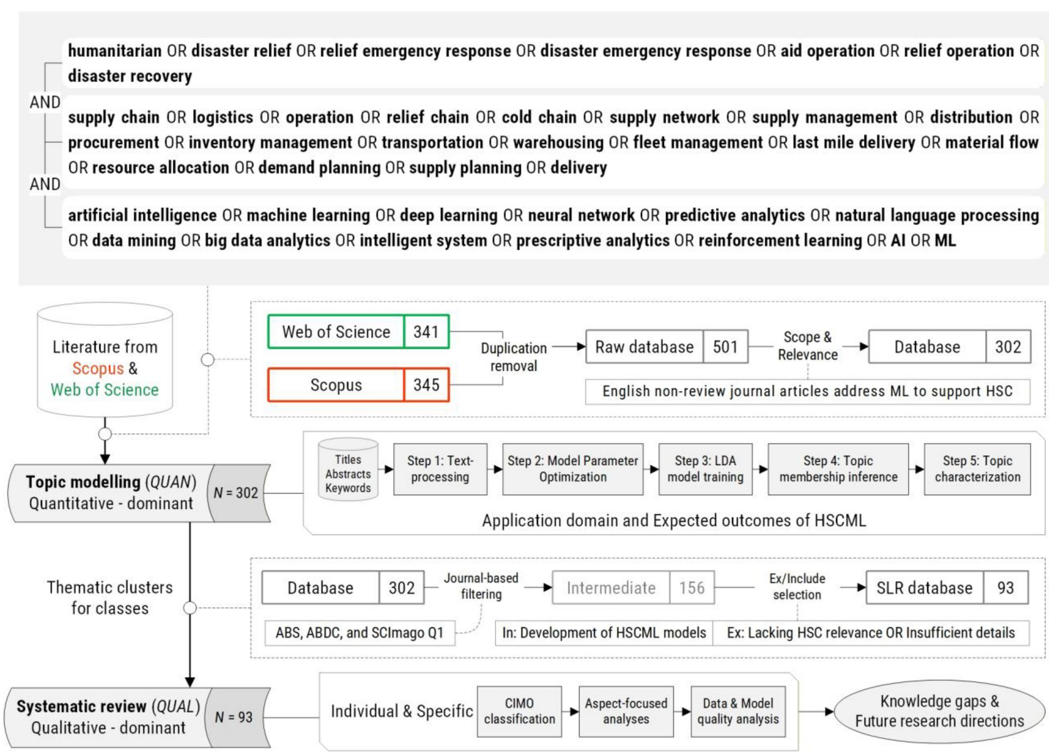
Based on the three targeted core elements: *humanitarian*, *supply chain management* and *ML*, a terminological expansion was carried out to develop a comprehensive three-field keyword structure detailed in Figure 1. The ML field keyword use generalised parent-category and technologically related terms, e.g. "deep learning", "neural network", "predictive analytics" and "prescriptive analytics" that encompass all specific technique variants and methodological overlaps, avoiding the risk of omitting novel or hybrid methods that technique-specific search terms might overlook. This structure was used to search titles, abstracts and keywords for English-language journal papers in Scopus and Web of Science (updated to 31 December 2025), with no publication date restrictions. This review excludes grey literature (e.g. technical reports, white papers) to ensure peer-review processes, reproducibility and technical specifications.

The papers' metadata were imported into *EndNote*. An aggregation with duplication removal yielded 501 papers. Abstracts were examined by two independent reviewers for relevance and originality (e.g. excluding papers addressing ML but not to support HSC, retraction and literature reviews). The data set was finalised at 302 journal articles. The inclusion and exclusion criteria, along with the number of papers after each screening step, are illustrated in Figure 1. To avoid missing relevant papers in emerging or specialised journals, the filtering based on journal ranking is placed only before the SLR phase.

2.2 Topic modelling

LDA was used to identify underlying thematic structures within the 302-paper corpus at the semantic level. LDA represents topics as probability distributions over words, while papers have probability distributions over topics (Zhao *et al.*, 2024b), thus enabling the direct examination of semantic content (i.e. title, keywords and abstract) when forming research clusters. The LDA models are governed by Dirichlet priors and built through an iterative training process. Inference is then conducted for each reference to determine its

Figure 1 The flow of the sequential mixed-methods systematic literature review



Source: Authors' own work

topic-probability mixture (Figure 1), with a membership threshold of ≥ 0.3 considered as significant topic membership. Each topic is uniquely described in two aspects. *Saliency* prioritises frequently occurring terms (Chuang *et al.*, 2012), whereas *relevance* ranks terms by their distinctiveness, emphasising word exclusivity (Sievert and Shirley, 2014).

This study used natural language processing (NLP) packages in Python: *NLTK* for text preprocessing, including tokenisation, stop-word removal and lemmatisation; *gensim* for LDA model implementation, topic extraction and n-gram detection; and *pyLDAvis* for term list extraction. Domain terminology standardisation and abbreviation handling were implemented through a custom thesaurus. In Step 2, different model configurations were tried and their performance compared. Title and keyword weights (1–5) relative to abstracts were applied. Weight combinations and topic numbers ranging from 5 to 20 were grid-searched, with the formation achieving the highest *topic coherence score* (C_v measures semantic coherence based on the co-occurrence patterns of top topic terms) selected for in-depth analysis and model parameters fixed (end of Step 3).

Following topic membership inference for each paper (Step 4), reviewers engaged qualitatively with the member papers and the top-ranked *saliency* and *relevance* terms for each topic in Step 5 (Figure 1), prioritising studies with higher topic membership probabilities as more exclusive and representative of the topic's thematic core. For each topic, reviewers identified the distinguishing ML application domain, characterised by its HSC purpose and operational focus, which constitutes the intervention (I) class in the CIMO framework. Concurrently, the improvements and claimed impacts reported in those papers, including the rationale for the ML application, constitute the outcome (O) classes. Where a first-level topic contained a disproportionately large number of studies and exhibited internal thematic heterogeneity upon qualitative review, another iteration of Step 2 – Step 4 was applied to that topic's corpus alone to identify finer-grained subtopics (i.e. Topics 1, 2, 3, 7 and 8). In total, 14 topics were characterised. Topics whose member studies exhibited a higher probability of belonging to other topics (Topics 6 and 8.2) were not characterised. Across the topics and subtopics, six I classes and five O classes were identified (Figures 2 and 3).

2.3 Systematic review

Metadata and full texts of 302 papers were further filtered based on the journals, following a multi-year combined list of ABS, ABDC and SCImago Q1 and the research focus on ML model development. Studies that solely focused on socio-technical aspects or lacked sufficient development details (e.g. data, model specifications) were excluded (Figure 1). Both criteria are methodologically required to ensure that RQs are answered by analyses on good (i.e. journal filtering) and complete and relevant data (i.e. model development filter). The final set of 93 papers was analysed in the systematic review.

Applied flexibly as Denyer and Tranfield (2009) intended, an adapted CIMO classification framework was used to characterise each HSCML study, i.e. in which context an ML application was deployed (C), what intervention was proposed (I), how it was technically developed through data and algorithms (M) and what outcomes were expected (O) (Figure 3). The C classes are adapted from previous HSC literature reviews (Chiappetta Jabbour *et al.*, 2017; de Camargo Fiorini *et al.*, 2022). The I and O classes are derived from LDA topic characterisation (see Section 2.2). For the M classes, while standard CIMO mechanisms describe human reasoning, motivation or social processes, ML intervention mechanisms are inherently technical. Data determines what patterns can be learnt, while algorithms determine how learning occurs (training) and how inputs are transformed into outputs (inference). Defining M as data and algorithms thus preserves the analytical separation between *what is done* (I) and *how it is done* (M), directly enabling the answers to *RQ2* and *RQ3*. Multiple class tags can be used for each paper.

Table 2 summarises the quality assurance practices applied throughout this review. It is worth noting that the analytical scope does not extend to specific developing pipeline practices (e.g. data processing, feature engineering and hyperparameter tuning). These context-, data- and modelling-purpose-specific investigations call for sub-field SLRs that could be facilitated using frameworks aggregated in this paper.

3. Systematic literature review results

Three reviews were conducted, covering multiple aspects of the selected 93 HSCML papers, following the C, I, M and O classes. Figure 4 illustrates the complete analytical flow, from the primary analytical focus of each subsection.

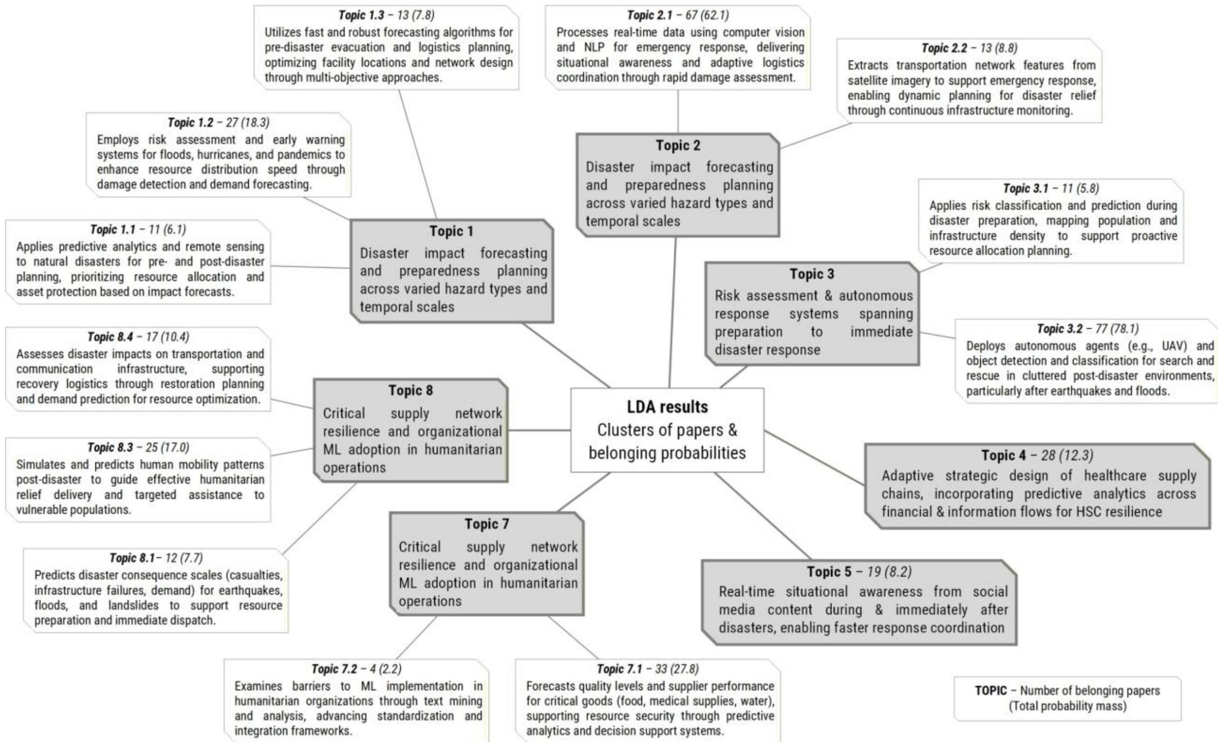
3.1 Context-focused analysis

The heatmaps in Figure 5 examine Context (C) × Intervention (I) × Outcome (O) intersections to address *RQ1*. The mechanism – M dimensions (data and algorithms) will be analysed in Sections 3.2 and 3.3, as problem requirements primarily drive technical choices.

Regarding country type and urbanisation, HSCML applications in developed countries and LMICs primarily focus on real-time situational awareness and logistics optimisation, based on predictive analytics (e.g. tree-based and deep learning) and prescriptive analytics (e.g. RL) [Figure 5(b)]. RL was used to ensure computational efficiency while maintaining optimality close to that of exhaustive optimisation methods (Pan *et al.*, 2025). HSCML in developed-country contexts tends to lean more towards critical infrastructure management, whereas in LMICs, it is more often focused on consequence prediction. *LDCs* are underrepresented in HSCML studies (5.7%), which are mostly set in rural contexts. HSCML in the context of *LDCs* uniquely prioritised disaster impact forecasting [e.g. mapping flood emergency aid requirements (Rahman *et al.*, 2021)], while continuous monitoring capabilities in these countries focused on pressing issues, such as water quality monitoring, war and conflict (Heuschmid *et al.*, 2025) [Figure 5(d)].

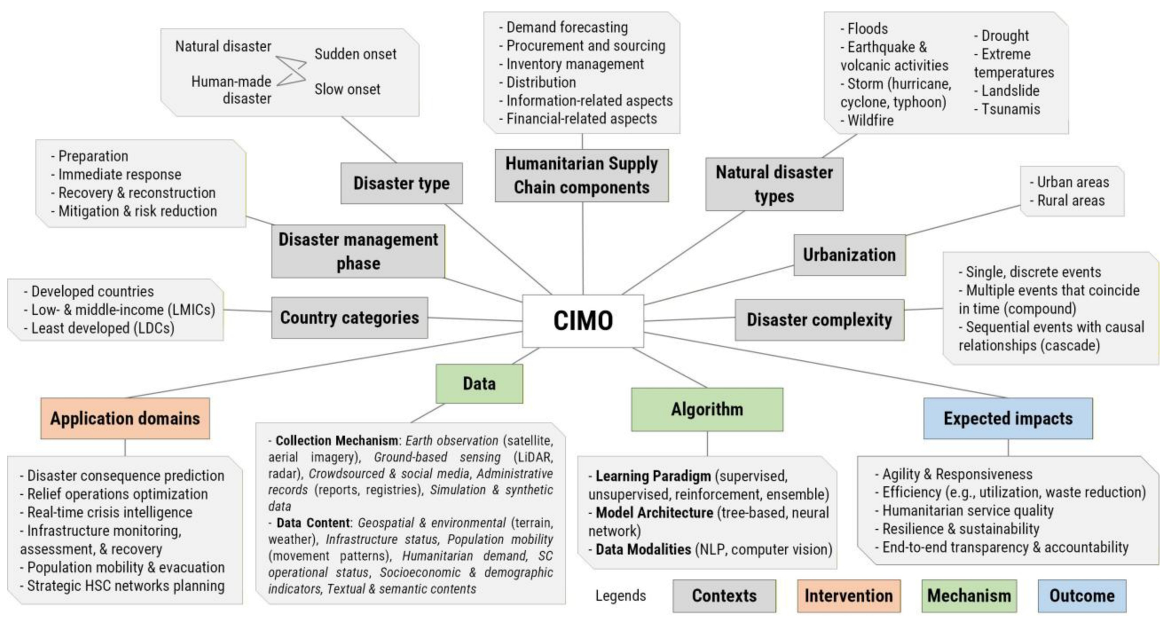
Natural disasters with sudden onset is the dominant focus, primarily addressing real-time situational awareness in response [e.g. UAV-based detections (Hazarika *et al.*, 2025)], while *natural disasters with slow onset*, such as droughts (Araya *et al.*, 2023) or pandemics (Bhullar *et al.*, 2022), demand resource optimisation with a focus on preparation [Figures 5(a) and (e)]. *Earthquakes and seismic* events are the most common, focusing on building damage assessment and monitoring

Figure 2 Summary of characterised topics based on LDA results



Source: Authors' own work

Figure 3 Classes in the applied CIMO classification framework



Source: Authors' own work

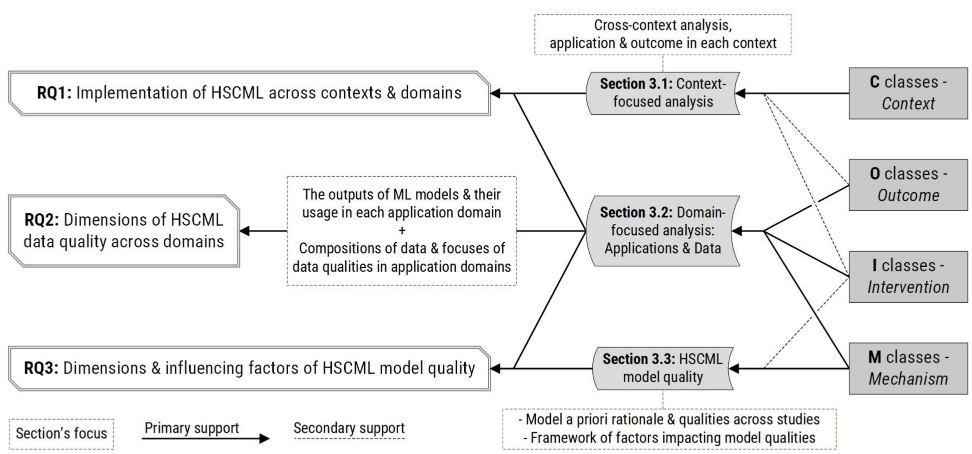
(Chen and Cho, 2022) and response coordination (Amin Amani *et al.*, 2025), with *tsunamis* addressed as cascading events requiring early warning, vulnerability assessment and evacuation (Song *et al.*, 2017; van Steenbergen *et al.*, 2023). *Flood and storm*, as well as *wildfire, studies* emphasise infrastructure monitoring (e.g. communication and power systems) and social media analytics for targeted humanitarian logistics. By contrast, *storm, flood, drought and landslide studies* feature risk management with occurrence and consequence prediction (e.g. groundwater salinity prediction).

Table 2 Methodological quality assurance considerations

Criterion	Practice	Implementation
Credibility	Review method design – RQs coherence	Sequential mixed-methods SLR design aligned with RQs (Figure 1); I, M and O classes in the CIMO framework are specifically adapted; each class serves a distinct analytical purpose mapped to the RQs (see Section 3)
	Validated empirical grounding of classes	I and O classes derived from LDA topic characterisation of the 302-paper corpus through qualitative sense-making, not imposed *a priori*. All 93 SLR papers were successfully assigned to these classes, confirming taxonomic stability
Transferability	Methodological transparency and availability	Full process documented step-by-step with rationale in Section 2 and Figure 1. Supplementary reviewed paper lists submitted for manuscript review; Python LDA implementation code available upon request
Dependability	Class theoretical grounding	C classes anchored in established HSC literature, ensuring contextual validity M classes were theoretically grounded as the ML mechanism enables interventions
	LDA parameter stability	Grid searches ensure the coverage of LDA model parameter; C_v coherence-metric-based model selection
Confirmability	Parallel qualitative processes	Two independent reviewers on each material-related process (abstract screening, full-text screening, LDA topic characterisation, CIMO tagging). Cohen's Kappa calculated across tasks: [min, max] of [0.50, 0.90], mean 0.73 and median 0.75 For the systematic review of 93 papers, thematic syntheses and interpretations were reviewed and validated by another researcher Deliberately resolve disagreements through a third non-reviewer researcher in biweekly meetings throughout the analysis process

Source(s): Authors' own work

Figure 4 Analyses in SLR based on the CIMO classification framework



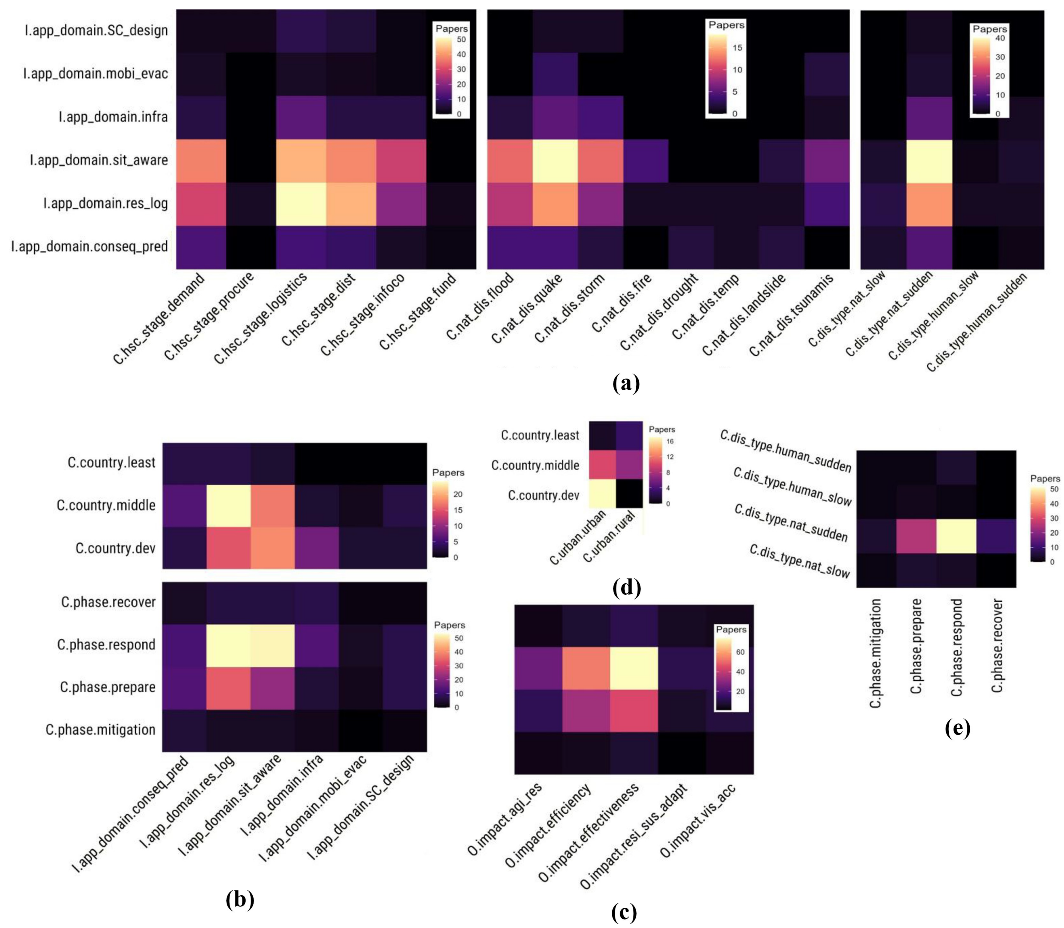
Source: Authors' own work

Human-made disasters with slow onset exclusively target resource optimisation, such as refugee resettlement and long-term aid planning (Ahani *et al.*, 2021). Meanwhile, *human-made disasters with sudden onset* prioritise situational awareness in war and military contexts, such as satellite imagery in monitoring humanitarian crises (Zhao *et al.*, 2024a), search and rescue in collapsed structures (Hu *et al.*, 2022) and anti-personnel demining (Heuschmid *et al.*, 2025). This is also the context for compound disasters, which necessitate adaptive policies, e.g. armed conflict, infrastructural collapse and service disruptions (Gong *et al.*, 2024; Zhao *et al.*, 2024a). The deployment of AI in conflict settings introduces distinct legal and ethical requirements, including compliance with proportionality obligations under International Humanitarian Law that current HSCML technical literature has yet to address (Dorsey, 2026).

Preparation phase prioritises strategic HSC design on pre-event allocation optimisation, such as inventory management and procurement planning (Quiliche *et al.*, 2023). The *response* phase focuses on real-time intelligence for rapid damage assessment and search and rescue operation (Hu *et al.*, 2022) and immediate resource distribution and procurement [e.g. multi-agent emergency response (Nadi and Edrisi, 2017)]. *Recovery* phase demonstrates higher emphasis on resilience and sustainability for long-term reconstruction [e.g. communication and transportation infrastructure monitoring and restoration (Ghannad *et al.*, 2021)] [Figure 5(c)]. The *mitigation* phase prioritises risk assessment and prevention through predictive likelihood and consequences insights [Figure 5(b)] [e.g. risk mapping (Rahman *et al.*, 2021), continuous infrastructure monitoring and contingency planning (Jana *et al.*, 2023) and essential supply resilience and sustainability (Zhao *et al.*, 2025)] [Figure 5(c)].

Apart from social media analytics, the *demand forecasting* stage used remote sensing to anticipate event consequences (e.g. building damage assessment for the prioritisation of

Figure 5 Heatmaps of sub-class C-I-O intersections indicating research focuses



Source: Authors' own work

evacuation (Xia *et al.*, 2023) or relief demand prediction based on dynamic population distribution (Lin *et al.*, 2020)) [Figure 5(a)]. ML in *distribution* emphasises real-time intelligence on humanitarian demands [e.g. NLP (Kiavash *et al.*, 2026)] to optimise precise and robust last-mile delivery [e.g. RL (van Steenbergen *et al.*, 2023; Peng *et al.*, 2025)]. HSCML of the *logistics* stage predominantly addresses transportation limitations in disrupted disaster zones [e.g. truck-UAV fleet deployment (van Steenbergen *et al.*, 2023)], rapid resource optimisation decision-making [e.g. RL for optimisation (Yu *et al.*, 2021)] and transportation networks monitoring and recovery (Jana *et al.*, 2023). *Procurement* remains underexplored, focusing on HSC design [e.g. optimal vendor selection under uncertainties (Hu *et al.*, 2024)]. Due to the lack of actual disaster data and the requirement for data completeness, synthetic data were widely used in these contexts. Apart from ML-integrated information systems for extracting HSC insights (e.g. NLP for social media pipelines), the information flow saw an emerging use case of re-establishing communication infrastructure, such as 5G networks (Yağcıoğlu, 2025). Financial aspects remain significantly underexplored, with a focus on socioeconomic-based resource allocation [e.g. the distribution of wealth and poverty (Chi *et al.*, 2022), humanitarian relief and stabilising food prices through national fiscal policies (Xiong *et al.*, 2024)].

3.2 Domain-focused analysis: applications and data

This section details the six application domains. Together with Section 3.1, this section directly addresses *RQ1* by characterising how HSCML has been operationalised and *RQ2* by reviewing the data quality dimensions relevant to each domain. The findings from this section also support the model quality dimensions and the influencing factors examined in Section 3.3 in addressing *RQ3*.

3.2.1 Disaster consequence prediction

This domain focuses on the predictive analytics of potential disasters' outcomes, including casualties, infrastructure damage, population displacement and resource requirements, focusing on earthquakes, landslides and floods (Terti *et al.*, 2019; Rahman *et al.*, 2021), providing crucial inputs for the preparation and mitigation phases. The targets of prediction could be modelled as either categorical (e.g. damage states, injury severity) or numerical (e.g. numbers of deaths and injuries) (Galera-Zarco and Floros, 2023; Biswas *et al.*, 2024; Zhang *et al.*, 2025).

Recorded data of the affected were critical in gaining insights into likelihood and severity, e.g. environmental and geophysical data to predict earthquake consequences, duration of precipitation to predict vehicle-related fatality and patients' records in predicting injury classes (Khosravi and Rabbani, 2026). Socioeconomic indicators are also involved both as the potential scale of vulnerability (e.g. house size, number of commuters) and the population's preparedness (e.g. number of emergency centres) (Terti *et al.*, 2019; Gong *et al.*, 2024; Zhang *et al.*, 2024a).

Regarding ML algorithms, tree-based ensemble methods are the most prevalent, with random forest and other gradient boosting methods achieving high accuracy (Terti *et al.*, 2019; Araya *et al.*, 2023; Gong *et al.*, 2024). Naturally, this domain is evolving towards multidimensional impact assessments, e.g. combining death and injury counts, infrastructure damage levels (Biswas *et al.*, 2024), cascading effects and the temporal evolution of impacts. These demands encourage the use of deep learning architectures, such as the sequence-to-sequence long short-term memory (LSTM) model for disaster needs prediction in different phases (Nguyen *et al.*, 2022). This domain also advanced from deterministic to probabilistic approaches, enabled by tree-based ensemble methods (e.g. random forest) to estimate the probability of groundwater salinity severity (Araya *et al.*, 2023).

Strong attention to data accuracy and credibility was observed through authoritative sources and cross-validation (Rahman *et al.*, 2021; Xiong *et al.*, 2024), or expert verification (Quiliche *et al.*, 2023). As the scopes of cause-effect are broad, data completeness and relevance were addressed through feature-engineering and data fusion, e.g. multi-year, multi-aspect, multi-source coverage (Rahman *et al.*, 2021; Biswas *et al.*, 2024). The class imbalance resulting from extreme events' rareness (specific to supervised classification ML) was addressed through sampling and stratification techniques (Araya *et al.*, 2023). Less explored are the domain-specific impacts of data temporal-spatial resolutions, data harmonisation standardisation (e.g. attribute definitions, fusion techniques) and factors that constitute data credibility/trustworthiness.

3.2.2 Relief operations optimisation

This domain focuses on resource allocation at the operational level during or immediately post-event, such as specific relief items based on population dynamics, seasonal and regional characteristics (Lin *et al.*, 2020). Apart from traditional multilayer perceptron (MLP) networks, more sophisticated approaches emerged to handle temporal dependencies and complex patterns, such as LSTM [e.g. fuel demand prediction (Fuqua and Hespeler, 2022)]. Allocation plans can also be optimised while balancing cost efficiency and service-level effectiveness across multiple affected areas and time periods [e.g. RL for relief items (Ahmad *et al.*, 2025) and medicines (Vanvuchelen *et al.*, 2024) distributions, dynamic vehicle routeing and scheduling (Hssini, 2025; Peng *et al.*, 2025) and team-based relief effort optimisation (Wu *et al.*, 2021)]. An advantage of RL is in overcoming model scaling issues to provide fast decision-making in catastrophic and complicated events (van Steenbergen *et al.*, 2023; Hssini, 2025).

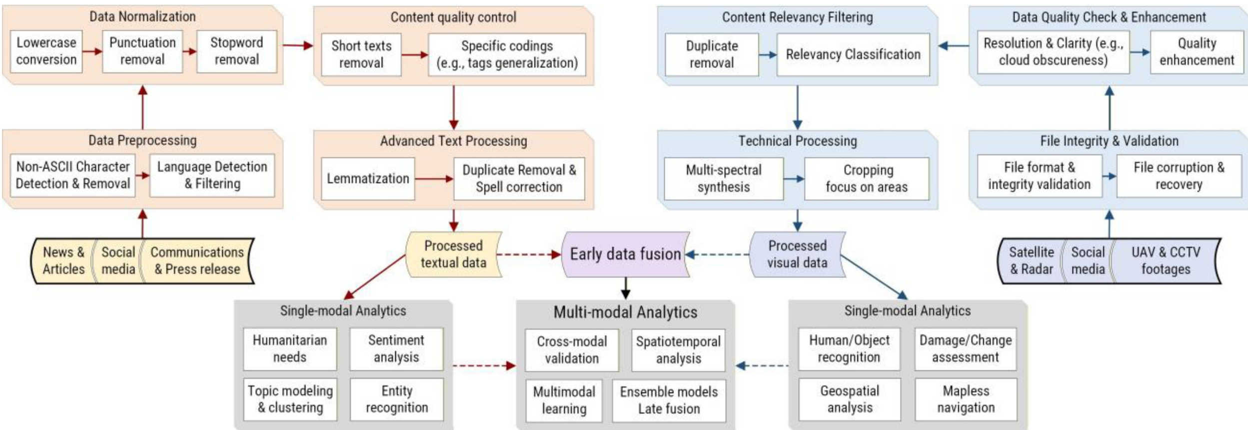
The primary inputs vary, including disaster parameters, recorded impacts and consequences, population demographics, resource availability and allocation/transportation costs and typically require data fusion. At this operational level, studies often use high-dimensional, high-resolution temporal and spatial data (Polushko *et al.*, 2024; Zhao *et al.*, 2024a), even real-time or near-real-time information, ranging from coordinates, velocity, direction, to agents' specifications and population density map (Lin *et al.*, 2020). Concurrently, the reliance on RL techniques necessitates synthetic data for a comprehensive, plausible and relevant representation of potential scenarios (Peng *et al.*, 2025), e.g. sequential action-reward interactions of disaster scenarios. Therefore, the realism of the scenarios used for RL learning was frequently mentioned. Domain expertise was used through either human-in-the-loop or ML-based verification [e.g. deep learning segmentation (Polushko *et al.*, 2024)], ensuring alignment with relevant theories (Lin *et al.*, 2020; Yu *et al.*, 2021; van Steenbergen *et al.*, 2023). Still, no framework was established to ensure data comprehensiveness and reliability, given the dubious robustness of deep learning techniques in extrapolation. Additionally, with the short timeframe for decision-making, potential biases were acknowledged in RL-recommended actions, especially with unreliable data as model inference inputs (Vanvuchelen *et al.*, 2024; Hssini, 2025).

3.2.3 Real-time crisis intelligence

This domain focuses on gaining insights and perceptions about the current emergency. Satellite images were used to assess the level of impact, especially in natural disasters, such as building damage (Hu *et al.*, 2023; Xia *et al.*, 2023; Zhang *et al.*, 2024b), while ground-penetrating radar can be used to map the situation of buildings (Hu *et al.*, 2022). On-site, convolutional neural network (CNN)-based computer vision studies concentrated on person detection in disaster scenarios, using data from cameras, thermal signals or LiDAR inputs from UAVs and CCTV to detect, navigate and locate people in need (Speth *et al.*, 2022; Wen and Chen, 2023; Diez-Tomillo *et al.*, 2024). Additionally, transformer- and LSTM-based NLP techniques can classify social media data into actionable categories, such as sub-events like emotional outbursts or local cases of disease transmission for relief and prevention measures [e.g. COVID-19 (Khatoon *et al.*, 2021; Gour *et al.*, 2022], or casualties, infrastructure damage and rescue progress in other disaster contexts (Alam *et al.*, 2019; Kumar and Singh, 2019; Fan *et al.*, 2020). A natural extension has been to develop multimodal deep learning models that combine CNN-based visual encoders and NLP-based encoders for text to rapidly assess situations with reduced uncertainty through cross-modal corroboration (Fan *et al.*, 2020; Garcia *et al.*, 2023; Zhang *et al.*, 2024b) (Figure 6).

Various techniques are used to ensure text data's completeness, including stemming, lemmatisation, word segmentation and translation expansion (Khatoon *et al.*, 2021; Ullah *et al.*, 2021; Bhullar *et al.*, 2022; Gour *et al.*, 2022). Apart from corroboration for credibility and filtering for relevance (Ullah *et al.*, 2021; Feng *et al.*, 2025), data accuracy was improved by preprocessing steps, such as spelling correction, phonetic substitution handling and denoising, super-resolution for image data (Fan *et al.*, 2020; Bhullar *et al.*, 2022; Gour *et al.*, 2022).

Figure 6 Multimodal textual and visual data processing pipeline



Source: Authors' own work based on Alam *et al*. (2019), Ullah *et al*. (2021), Zhang *et al*. (2024b)

However, the complications of multi-source data processing pipelines could introduce additional complexity and, in turn, uncertainties. These pipelines are rarely transparent, follow established frameworks or are well-reasoned based on the data sets at hand, obscuring the impacts on data quality dimensions. For example, inconsistent damage assessment scales and ambiguous definitions of "disaster-related" content affect data representativeness and completeness. Additionally, data are costly to label and validate (Fuqua and Hespeler, 2022). The scarcity of training data for deep learning models may also cause latent issues, including feature distortion, shortcut and non-coverage biases and domain shift, which are acute for CNN and transformer architectures (Zhu and Wang, 2023). Finally, despite the significance of low-latency high-resolution data in this domain, the deliberate trade-off between timeliness and accuracy in modelling (Speth *et al.*, 2022) has not been systematically addressed. These data quality concerns are likely the culprit for the significant mismatches observed by Kiavash *et al.* (2026) between social media-extracted demand signals and other reports.

3.2.4 Infrastructure monitoring, assessment and recovery

This domain supports physical damage and resilience assessment and the restoration of critical infrastructure, ensuring the continuity of essential services and responses. At the tactical level, deep learning techniques were used extensively to investigate the functionality and reliability of the logistics network, i.e. road and bridge systems and prioritising critical components to recover [e.g. radar-based structural assessment (Hu *et al.*, 2022), graph neural networks for network topology (Jana *et al.*, 2023)]. Damages to buildings' structural integrity and safety were also detected and assessed (Galera-Zarco and Floros, 2023; Zhao *et al.*, 2024a). Multi-agent RL models were proposed to coordinate UAV swarms for communication and computing infrastructure recovery (Faraci *et al.*, 2023).

HSC prioritises functionality in the objective function's components [e.g. call completion and service availability over communication quality and energy efficiency (Almalki and Angelides, 2019)] while keeping the feasibility of the solutions through the authenticity of scenarios [e.g., GenAI (Hazarika *et al.*, 2025)] and data realism [e.g. visual Turing test (Hu *et al.*, 2022), data abstraction quality (Jana *et al.*, 2023)]. This domain shows attention to data completeness, particularly in infrastructure data, e.g. heterogeneity through merging video footage and physics-based simulation (Galera-Zarco and Floros, 2023; Hu *et al.*, 2023; Zhao *et al.*, 2024a). A typical synthetic data pipeline includes foundation data acquisition, parameter space definition, synthetic data generation and realism enhancement (e.g. generative AI) (Hu *et al.*, 2022; Galera-Zarco and Floros, 2023).

However, presumptions regarding critical data configuration were not always strongly addressed, e.g. anomaly detection assuming cracked points are minor, with arbitrary thresholds (Chen and Cho, 2022). Frequently, data processing decisions were reasoned concerning individual quality regardless of how they might negatively affect other dimensions [e.g. deliberate annotation bias inconsistency (Zhao *et al.*, 2024a) and realistic scenarios by Generative AI may not be comprehensive (Hazarika *et al.*, 2025)].

3.2.5 Population mobility and evacuation

This domain focuses on the movement pattern of population during and after events, which is valuable for HSC operation optimisation at the tactical and operational levels [e.g. predictions of travel and service times for evacuation using XGBoost (Nabavi *et al.*, 2022) and emergency behaviour and emergency mobility using deep RL (Song *et al.*, 2017)]. Due to the domain's dynamic nature and its relationships with different social and economic factors, data processing heavily involved advanced location discovery and behavioural analytics, e.g. location's functions and familiarity and social relationships (Song *et al.*, 2017). The challenges were acknowledged in fusing positioning, event and transportation network data, news reports and feature engineering to ensure data relevance and completeness [e.g. transportation mode with destination labelling for millions of anonymised users for several years with billions of GPS records (Song *et al.*, 2017)]. Validating data quality is often overlooked, especially in complex data fusion processes, despite its systematic implications for model quality. Another issue is the practice of assessing data quality using model accuracy indicators, an indirect approach that may create false expectations about the model's performance in real-world operations.

3.2.6 Strategic humanitarian supply chain networks planning

This domain optimises HSC facility decisions and network planning to enhance resilience. For example, network optimality can be estimated using topological measurements (e.g. centrality indices). A classification model trained on that pattern can quickly and efficiently optimise distribution centres' locations [e.g. deep neural network (Taouktsis and Zikopoulos, 2024)]. Similarly, the importance of facilities can be used to strategise resource deployment to precisely improve system resilience (Jana *et al.*, 2023). Models can predict operational performance across different scenarios and management strategies [e.g. cumulative demand, initial inventory, management strategies and cumulative delivered supply to predict the average cumulative delivery percentage and the daily cumulative percentage of fulfilled demand (Zhang *et al.*, 2024a)].

These studies require both synthetic and authentic data to ensure completeness, a manageable computational workload and model generalisability. For example, clustering was used to effectively and efficiently stratify disaster scenarios (Hu *et al.*, 2024). ML-based surrogate modelling was used to efficiently approximate the results of computationally heavy and repetitive programming tasks in optimisation [e.g. two-stage optimisation of pre-event location and resource allocation and post-event procurement, distribution and inventory (Pan *et al.*, 2025)]. Data credibility was classified into three levels, with historical data ranked the highest – used to validate the simulation models, which are then used to validate ML model results (Zhang *et al.*, 2024a).

While data generation parameters are mostly transparent, their rationale and verification are not always clear, affecting the basis for validity and realism (e.g. synthetic coordinates, distances between distribution centres and impact/consequence levels for each synthetic scenario). Some studies relied on experts for plausibility and representativeness verification (Zhang *et al.*, 2024a). However, data representativeness from an expert's perspective can be ambiguous regarding statistical distributions, completeness, or "operational possibility". Additionally, while data imbalance was frequently mentioned as a quality-related issue [e.g. nodes being the minority among many other nodes (Taouktsis and Zikopoulos, 2024)], the imbalance as a nature of the phenomenon is not inherently a data quality issue and should be considered under the appropriateness of modelling techniques.

The key HSCML data and ten data quality dimensions derived from analyses across different application domains are aggregated in Figure 7. It is noteworthy that operationalising these dimensions requires domain-specific indicators and thresholds (see Section 4).

3.3 Humanitarian supply chain machine learning model quality

Model quality considerations can be categorised into *a priori rationales* (i.e. reasoning for algorithms and pre-trained models) and *post-training testing and evaluation*. This separation not only provides insights into the dimensions of model quality and the technical considerations in the modelling process but also helps identify quality gaps in model evaluation.

3.3.1 A priori rationales for machine learning techniques

Data-related rationales address only partially the data-related challenges in the humanitarian context (see Section 3.2). Apart from big data analytics, handling noisy data became critical for NLP and transformer-based models working with unofficial, unvalidated sources, such as real-time handling of colloquial expressions, abbreviations and multilingual social media posts (Kumar and Singh, 2019; Saleem *et al.*, 2024) (Figure 8). To gain preprocessing insights into data sets, data exploration is usually deployed, in which unsupervised ML can be valuable [e.g. LDA (Zhao *et al.*, 2024b)].

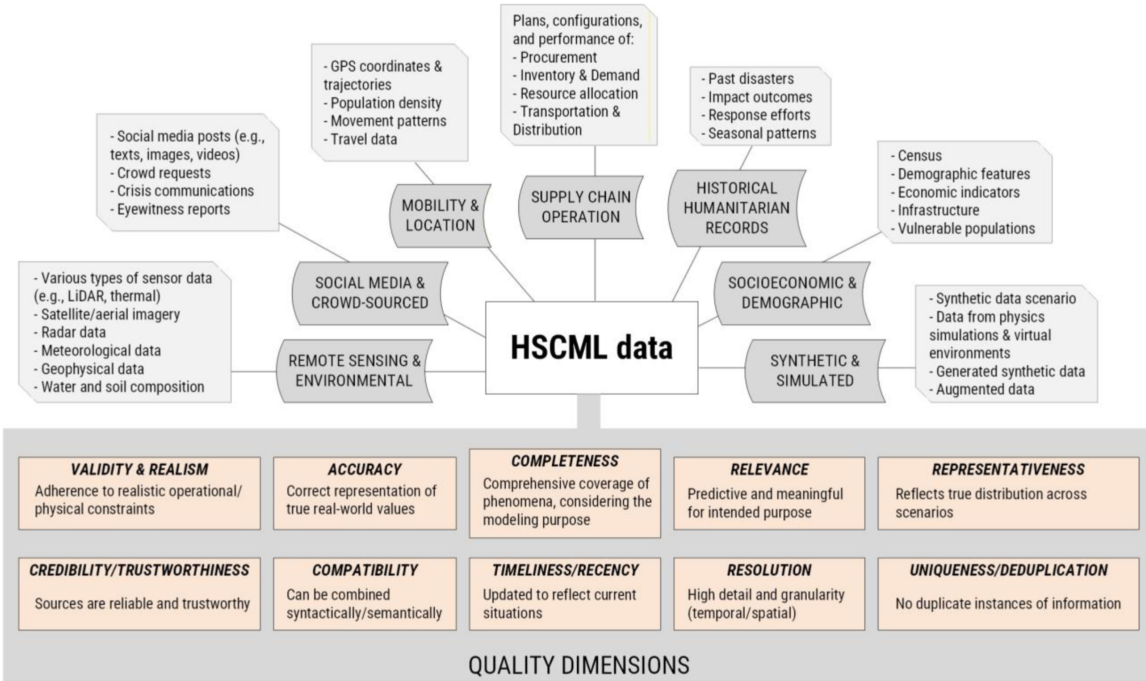
Some rationales are specific and decisive towards methodological choice. For example, the rarity of disaster events can be addressed through transfer learning (Zheng *et al.*, 2022), or data imbalance and scarcity are addressed by stratification techniques (Hu *et al.*, 2024). Additionally, purpose-built architectures and pre-trained models have proven critical in complex HSCML solutions, such as object recognition and image segmentation and reconstruction [e.g. YOLOv5 for building damage (Hu *et al.*, 2023), segmentation models (Polushko *et al.*, 2024)]. Handling complex multi-factor interactions and cascading effects is another popular rationale, such as to handle temporally dependent data [e.g. LSTM for demand forecasting (Nguyen *et al.*, 2022)] and complex multi-objective optimisation of spatial-temporal decisions under uncertainty [e.g. multi-agent RL for UAV coordination (Hazarika *et al.*, 2025)].

Modelling technique rationales focus on operational and deployment requirements. Modelling accuracy remains paramount in HSCML, e.g. resource misallocation can be catastrophic, motivating the choice based on known theoretical performance [e.g. deep learning (Song *et al.*, 2017), tree-based and boosting algorithms (Nabavi *et al.*, 2022), stacking ensembles (Zhang *et al.*, 2025)]. Computational efficiency in real-time inference and training/retraining addresses the demands for rapid predictions and model updates in emergency responses [e.g. surrogate modelling for relief logistics (Pan *et al.*, 2025), CNN for victim detection (Wen and Chen, 2023)], in which model size is a deliberate trade-off between accuracy and deployability (Hu *et al.*, 2023; Heuschmid *et al.*, 2025).

Another trade-off is between interpretability and other model quality dimensions, such as explainability, reliability [e.g. tree-based algorithms for transparent and deterministic feature importance (Araya *et al.*, 2023; Quiliche *et al.*, 2023)]. Apart from the advantage of ML in its non-assumption of predictor-target relationships, model robustness is cited as critical to ensure performance when dealing with outliers and data uncertainties, e.g. robustness to viewpoint variations and illumination conditions (Wen and Chen, 2023; Polushko *et al.*, 2024; Zhao *et al.*, 2024a) and seasonality (Vanvuchelen *et al.*, 2024). Finally, the method/technique's proven track record is deemed significant, citing the fit with data and prior studies (Garcia *et al.*, 2023; Yağcıoğlu, 2025).

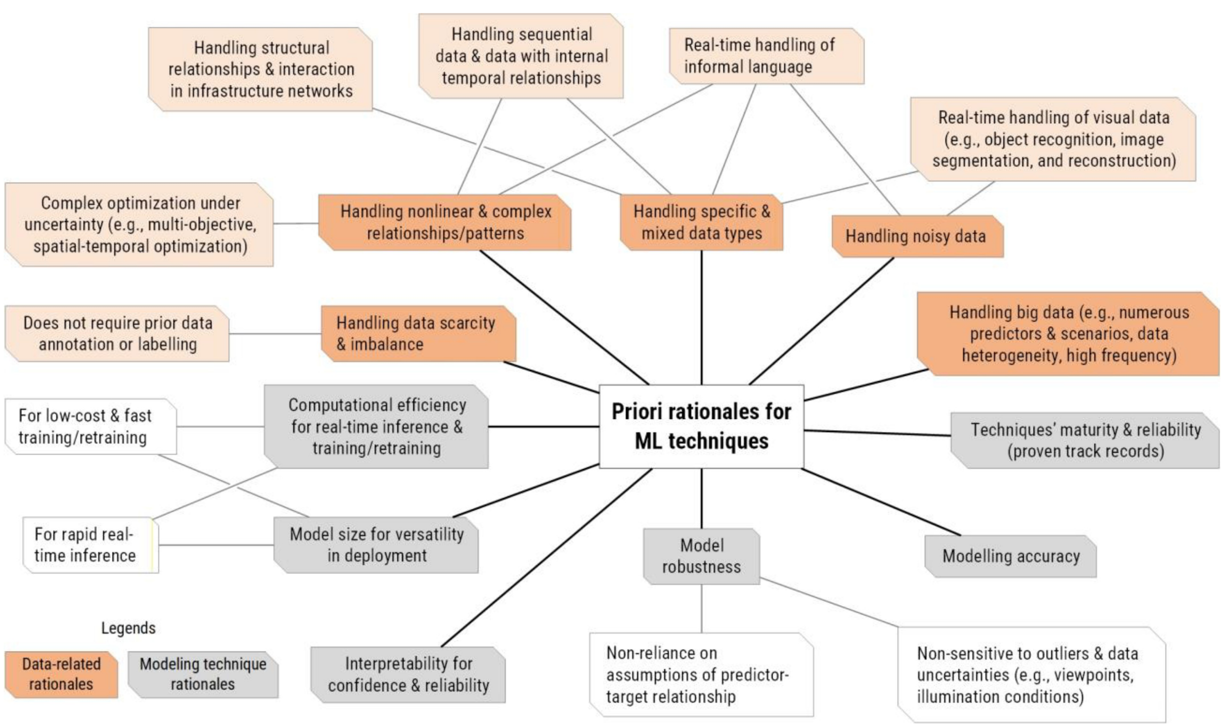
However, significant gaps exist. Firstly, there is a potential mismatch between the rationales' generality and techniques and algorithms' specificity, i.e. studies reasoned technical choices (e.g. neural networks, or pre-trained models) based on ML's umbrella characteristics, such as nonlinear relationships and big data handling. Studies used vague, ambiguous and

Figure 7 Data and data quality framework for HSCML



Source: Authors' own work

Figure 8 Aggregated a priori rationales of HSCML methodology consideration



Source: Authors' own work

superficial reasons, such as "high-performance", "high accuracy", or even "widely used". These weaknesses in the methodological basis and premature conclusions could lead to overlooking suitable methodologies, especially given the trade-offs among model quality dimensions and the strong individuality of implementations.

Secondly, despite the growing recognition of uncertainty quantification in ML (Abdar *et al.*, 2021), HSCML studies provide limited rationale for handling and expressing uncertainty beyond data-related robustness (17/93 reviewed articles ~ 18.3%). The lack of uncertainty-conscious methodological choices (e.g. confidence and prediction intervals, noise decomposition, quantile predictions, ensemble variance) might affect model applicability, as both aleatoric and epistemic uncertainties can propagate through HSC, from strategic to operational levels.

Thirdly, despite explainable AI has become recognised as essential for high-stakes decision-making (Barredo Arrieta *et al.*, 2020) and even a structural requirement for humanitarian AI (Dorsey, 2026), there is still limited consideration of explainability beyond feature importance. In HSC's collaborative environment, stakeholder-centred explainability is non-negotiable for execution. More critical decisions favour the use of tractable calculations and situation-solution relationships at the expense of model accuracy (e.g. tree-based models rather than neural networks).

Fourthly, despite the criticality of fair and equitable ML (Caton and Haas, 2024), efforts in HSCML to incorporate and consider these qualities in solution design are still very limited (11/93 reviewed articles ~ 11.8%), e.g. resource allocation with fairness-integrated objective functions (Vanvuchelen *et al.*, 2024; Ahmad *et al.*, 2025). This negligence indicates a disconnection from the technical choices and the desired model quality.

Overall, in comparison with well-known, established general ML system design frameworks (Sculley *et al.*, 2015), HSCML is currently behind in some key areas, including overall ML-based solution design, interpretability and conscious and predetermined trade-offs between aspects of model performance (e.g. errors' costs) and equity and fairness. These shortcomings highlight the urgency of a framework for model quality dimensions.

3.3.2 Model quality dimensions and influencing factors

The identified HSCML model quality dimensions can be categorised into three groups (Figure 9). *Technical performance* dimensions were examined beyond mere accuracy. Computational efficiency and responsiveness address time-critical humanitarian operations through inference latency measurements or initialisation time requirements [e.g. CNN-based models (Diez-Tomillo *et al.*, 2024)] and training duration constraints [e.g. transformer architectures (Saleem *et al.*, 2024)]. Robustness and generalisability ensure reliable performance through training monitoring with [e.g. early stopping to control overfitting and model bias (Hu *et al.*, 2023), accuracy under imperfect data (Gong *et al.*, 2024) and cross-validation across (Jana *et al.*, 2023)].

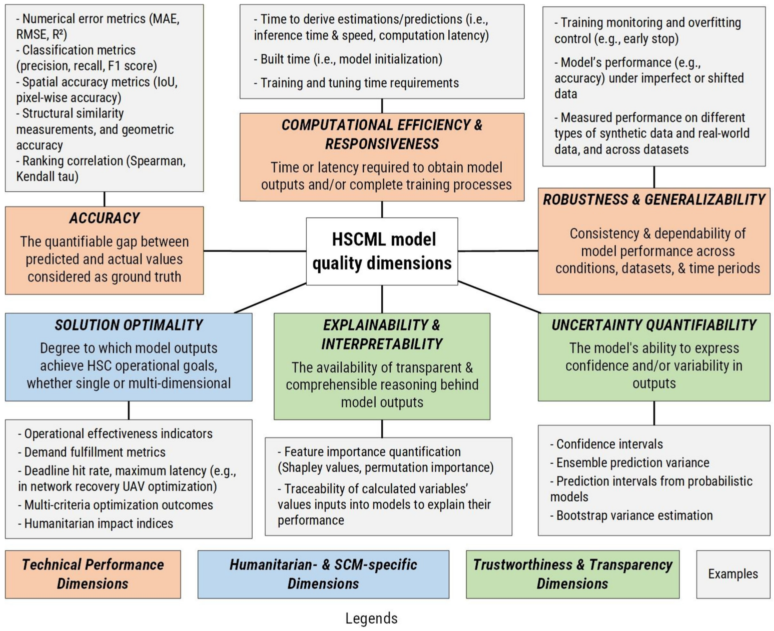
Regarding *trustworthiness and transparency*, a similar result to Section 3.3.1 was derived, as explainability and interpretability are conflated with feature importance quantification [e.g. Shapley-based importance (Feng *et al.*, 2025), permutation importance (Terti *et al.*, 2019)] and calculation traceability [e.g. to explain HSC's predicted performance aspects (Zhang *et al.*, 2024a)]. Although not always included, *uncertainty* was quantified through different techniques, such as the variance of results from ensembles and bootstrap samples (Ahani *et al.*, 2021; De Santi *et al.*, 2021). Regarding operational performance, *solution optimality* evaluates the objective achievement of HSCML-optimised outputs through indicators [e.g. survivability metrics (Lee and Lee, 2021)], demand fulfilment metrics (Vanvuchelen *et al.*, 2024), multi-criteria performance trade-offs, such as efficiency and equity (Ahmad *et al.*, 2025) and flow-specific metrics, such as energy consumption and delay in network recovery (Yağcıoğlu, 2025).

Regarding fairness and equity, the disaster consequence prediction domain demonstrates emerging but limited considerations, primarily through socioeconomic input variables (e.g. cost of living, per capita income) and a vulnerability-centred model design (Biswas *et al.*, 2024). Cost of errors (e.g. deprivation cost of misclassifying at-risk households) can be used to deliberately adjust the balance between accuracy and humanitarian optimality (Quiliche *et al.*, 2023). Infrastructure monitoring, assessment and recovery rarely engage with fairness and equity. In RL-based optimisation, considerations can be injected into the reward function [e.g. social vulnerability based on income, education and housing conditions (Ghannad *et al.*, 2021) and fairness index for equitable resource allocation (Yağcıoğlu, 2025)]. Similarly, some relief operation optimisation studies balanced efficiency, effectiveness and equity through deprivation costs [e.g. penalty costs to ensure no survivors are severely disadvantaged) (Yu *et al.*, 2021; Ahmad *et al.*, 2025]. Other studies addressed service equity for disadvantaged communities, such as unserved area penalties (Peng *et al.*, 2025) and geographic fairness in medicine access by minimising variance across facilities (Vanvuchelen *et al.*, 2024). Real-time crisis intelligence acknowledged geographic biases, such as social media attention is not uniformly distributed across disaster-affected areas (Fan *et al.*, 2020). Studies in strategic HSC network planning and population mobility and evacuation domains demonstrated a lack of consideration for fairness and equity, prioritising only technical efficiency (e.g. network efficiency, traditional supply chain performance metrics, prediction accuracy, evacuation time and costs).

Across domains, fairness and equity remain critically underexamined and are often considered as *post hoc* assessments rather than predetermined objectives. Both real-time crisis intelligence and disaster consequence prediction recognise that data availability and attention are geographically uneven, yet no corrective frameworks or systematic fairness integration have been proposed, thereby perpetuating spatial inequities and marginalisation (e.g. digitally disconnected populations). The balance among efficiency, effectiveness and equity is adjustable, but the quantification of weights, influenced by political, societal and cultural factors, was overlooked. In HSC networks, planning and infrastructure management, facility location and network design decisions can compound social disparities (e.g. transportation access, socioeconomic constraints and vulnerable populations' evacuation capacity), calling for more dedicated efforts in demographic disaggregation to ensure tailor-made yet equitable assistance.

Additionally, the existing dimensions focus primarily on static, lab-environment metrics rather than situational adaptability and application-oriented conditions, such as reliability in handling data imperfections and concept drifts (e.g. population dynamics, evolving conflicts, climate change) (Nguyen *et al.*, 2023). While transfer learning can enable generalisability, HSCML's ability to correctly render fundamental understandings was not evaluated, undermining

Figure 9 Aggregated framework for HSCML model quality dimensions



Source: Authors' own work

adoption confidence. Moreover, there is a lack of discussion of the importance of model quality dimensions and how their trade-offs should be considered. Finally, uncertainty quantification is not widely and systematically applied (e.g. aleatoric and epistemic), which could lead to ineffective and inefficient decision-making.

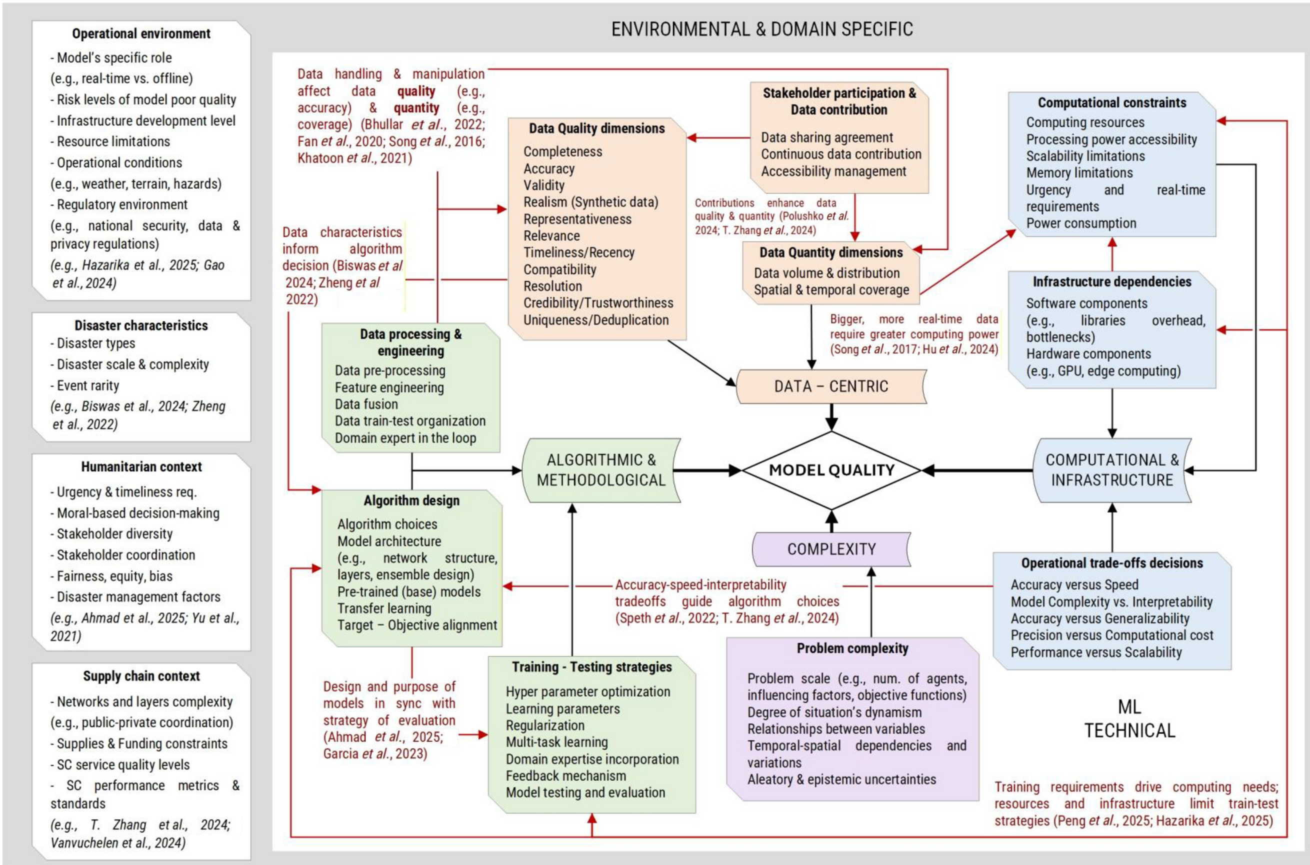
Figure 10 presents a comprehensive taxonomy of factors influencing HSCML model quality, thematically aggregated from the reviewed papers. Model quality is directly shaped by data-centric, computational and infrastructure, algorithmic and methodological and complexity factors. The outermost layer captures environmental and domain-specific influences, where decisions should be grounded in legal, ethical and situational awareness (Dorsey, 2026). This framework guides a systematic consideration of the factors that influence ML model quality during planning, development and improvement. Considering the recent development of ML in LSCM (Vlachos and Reddy, 2025; Nguyen *et al.*, 2026), there is a gap in the sustainability of HSCML models in deployment, i.e. functional prototypes are evaluated and tested locally, while regimes for HSCML application-oriented testing are nonexistent, overlooking critical elements (e.g. performance monitoring and retraining). In addition, HSCML relies on software, hardware and infrastructure that may be subject to bottlenecks and dependencies. Examples in the software realm include data supply chain and software dependencies (e.g. data quality dimensions along the pipeline, computational overheads and biases within base and foundational models). The complexity of solutions is even higher for multi-model chaining, coupling and stacking, but these were not discussed.

4. Discussions and future research directions

HSCML efforts were unevenly distributed across disaster types, management phases and geographic regions. There is limited attention to slow-onset events (e.g. droughts, pandemics) and human-made crises (e.g. conflicts and industrial accidents), despite their complex and emerging impacts. This review confirms the role of digital innovations, here HSCML, in the response phase (Maric *et al.*, 2022; Altay *et al.*, 2023) and indicates a lack of attention to recovery and mitigation, as well as strategic-level decisions on HSC sustainability and resilience. Compound and cascading disasters pose modelling challenges that have not been adequately addressed, representing opportunities for HSCML applications. HSCML research also neglected LDCs despite their humanitarian challenges. While human factors such as know-how and competency and technological constraints contribute to the barriers (de Camargo Fiorini *et al.*, 2022; Heuschmid *et al.*, 2025; Behl *et al.*, 2026), data scarcity is the most self-reinforcing obstruction across the ecosystem. The current sparse and fragmented historical disaster records preclude the applicability of ML to more complex, gradual and latent phenomena, further limiting our understanding of long-term sustainability and performance of HSCML systems and calling for cross-country data sharing, funded regional databanks and local expertise development to enable and generalise HSCML research.

Across the six HSCML application domains, two model architectures dominate: tree-based ensemble methods (e.g. random forest, XGBoost) and neural-network-based methods (e.g. CNN, LSTM, transformer). In the disaster consequence prediction domain, where tabular environmental and socioeconomic data predominate, *tree-based methods* prevail, offering native probabilistic outputs, interpretable feature importance and reliable performance. However, they are critically limited in extrapolation, i.e. unable to generalise beyond the training distributions, thereby constraining their applicability to significant and rare events. *Deep learning*

Figure 10 Summarised factors affecting HSCML model quality from the literature



Source: Authors' own work

architectures dominate high-dimensional, multimodal domains, including CNNs and graph neural networks in crisis intelligence and infrastructure monitoring for visual and spatial assessment, LSTMs for temporal modelling in demand forecasting and transformer-based encoders for NLP and multimodal pipelines. Despite their versatility, they are vulnerable to shortcut learning and feature distortion under data scarcity, require explicit uncertainty quantification methods and depend on *post hoc* approaches for explainability and fairness. *RL*, most frequently implemented through neural network architectures, has been mainly applied in distribution optimisation and infrastructure recovery contexts but carries its own quality considerations, such as reward-signal sensitivity, out-of-distribution state failures and limited explainability across decision trajectories.

In practice, deployable HSCML systems will increasingly involve multi-model pipelines where architectures are chained sequentially, e.g. transformer-based NLP, such as BERTopic, maps the demand feeding RL-based distribution optimisation and GNN-based network criticality coordinates with multi-agent RL for infrastructure recovery. Methodological decisions must therefore be considered at the pipeline (system) level rather than for each sub-problem in isolation, where architectural compatibility, propagation of failure modes and system-wide computational constraints collectively determine operational feasibility. The data quality framework (Figure 7), model quality dimensions (Figure 9) and influencing factors taxonomy (Figure 10) provide the reference for this system-level design, ensuring quality assurance is built into the pipeline architecture rather than assessed *post hoc*. This review also expands Kondraganti *et al.* (2022) by specifying six additional data categories beyond social media and clarifying their roles across application domains.

In response to the calls of Anjomshoae *et al.* (2022) and de Camargo Fiorini *et al.* (2022) for data-driven knowledge for humanitarian operations, four operational necessities for sustainable HSCML are recommended. *Firstly*, regarding the lack of good-quality data (Anjomshoae *et al.*, 2022; Kondraganti *et al.*, 2022), HSCML studies address data quality concerns piecemeal, focusing on data set-specific challenges and conforming to common practices without a cross-domain standard. This narrowness and individuality are critical, as they cause field-wide inconsistencies that go unaddressed and unresolvable. The aggregated data quality dimensions framework (Figure 7) addresses this by offering a comprehensive, domain-transferable reference for data quality.

Secondly, while ML-based systems need retraining (Vlachos and Reddy, 2025), data processing pipelines in HSCML are rarely transparent or well-grounded and regimes for continuous data quality assurance (e.g. pipeline maintenance) are nonexistent. As found in this review, HSCML studies are predominantly proof-of-concept evaluated in controlled settings. While retraining, monitoring and pipeline maintenance fall outside the typical design scope, they are critical gaps for long-term system sustainability in resource-constrained environments (Ülkü *et al.*, 2024). This confirms Maric *et al.*'s (2022) assessment that HSCML remains primarily technology-centric rather than humanitarian- and supply-chain-integrated, for which tensions with operational readiness have been observed (Dorsey, 2026).

This review indicates critical prerequisites for deployable HSCML, including data source integration and compatibility, pipeline-level quality management and broader data and model quality dimensions (Figures 7 and 9), calling for a comprehensive framework of HSC performance indicators, which could be built on the established LSCM foundation (Anjomshoae *et al.*, 2022). Data processing pipelines and modelling technique selection should be designed with purpose and tailored to contexts, rather than replicated from prior studies, given the individuality of data-model configurations and evolving data quality landscapes, e.g. improving sensor quality and deteriorating social media data caused by chatbots and generative AI content.

Thirdly, despite HSCML's reliance on multi-source data fusion, stakeholder participation and data governance factors (Figure 10) remain largely overlooked. HSCML data is inherently cross-organisational, requiring separate access and sharing agreements, yet the reviewed literature predominantly treats data as pre-given, bypassing the governance required to secure, integrate and maintain it. Given the identified barriers of resistance due to resource competition and accountability concerns (Anjomshoae *et al.*, 2022) and the prerequisite of data privacy, transparency and data management for stakeholder trust (Vlachos and Reddy, 2025), future research on governance, such as data valuation based on contributions to model performance, provenance tracking, cross-organisational quality assurance will be critical.

Fourthly, domain expertise through human-in-the-loop or expert verification is proven critical (de Camargo Fiorini *et al.*, 2022). As model complexity scales, the expertise required to validate model behaviours and outputs grows accordingly. Yet, this capability has not been nurtured and scaled proportionately, preventing the realisation of HSCML's true potential. It is known that automation bias and cognitive offloading can progressively erode judgement operators' independent judgement and oversight capacity through over-reliance on AI outputs, i.e. "cognitive shifting" (Behl *et al.*, 2026; Dorsey, 2026). Explainability is therefore the functional prerequisite for an operationally ready HSCML, enabling operators to interrogate model decisions, identify failures and maintain systems throughout their operational life cycle.

The review results indicate dominance of technical-induced model quality dimensions, both in *a priori* rationales and in the evaluation of model quality. At the same time, HSC characteristics, such as fairness and equity, are underexamined. Where present, the considerations are implemented narrowly through socioeconomic input variables and RL reward functions and treated as *post hoc* assessments. Incorporating fairness adds complexity to model design and may require trade-offs, as the framework in Figure 10 indicates, without improving the more obvious, easy-to-defend accuracy and computational metrics. Therefore, it is systematically deprioritised. This finding extends the concern by Nawazish *et al.* (2023) that the social pillar of sustainability in HSC remains nascent. It underscores the difficulty of translating humanitarian principles, such as impartiality, into measurable performance indicators. Future research should integrate equity throughout the HSCML development lifecycle, e.g. incorporating related indices as input features and model constraints, implementing demographic disaggregation in model outputs, developing tuneable multi-objective efficiency-equity trade-off frameworks, correcting geographic and demographic biases and establishing participatory design protocols that involve beneficiaries in the co-design of equity metrics. A view beyond social equity towards a full quadruple bottom-line sustainability framework, including economic, environmental, social and cultural elements (Ülkü *et al.*, 2024), will also benefit adoption, as environmental and cultural sustainability remain largely underexplored in the current HSCML literature.

Lastly, the review reveals persistent weaknesses in methodological rigour and operational readiness, i.e. a tendency to justify methodological choices by appealing to ML's umbrella characteristics rather than to contextual suitability and the specificity of algorithms/techniques. While concurring with Vlachos and Reddy (2025) on the versatility of neural networks, this review identified a multidimensional range of factors influencing model quality (Figure 10) that do not always favour deep learning. These loose reasonings and the limited consideration of trade-offs between model quality dimensions indicate again that HSCML publications' focus remains on proof-of-concept rather than method optimality (Sculley *et al.*, 2015; Maric *et al.*, 2022; Altay *et al.*, 2023). HSCML spans across HSC and computer science communities with different evaluation standards and has inherited the latter's benchmark-oriented confidence without the former's insistence on quality control. Integration and system-scale testing and accountability frameworks with regimes for performance monitoring and retraining (e.g. Nguyen *et al.* (2026)) should be developed to facilitate HSCML deployment.

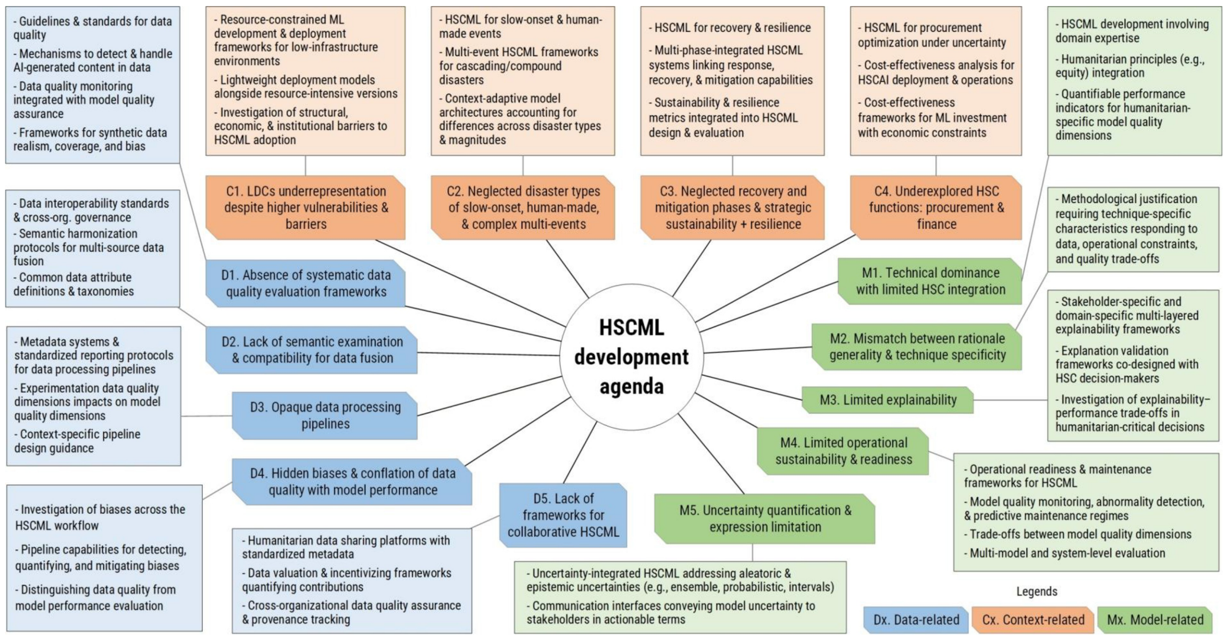
The discussions above highlighted the gaps that hinder HSCML from maturing from technical prototypes to impact-oriented, operationally ready applications. These discussions are not only enabled by but also operationalise the three synthesised frameworks, providing a principled reference for development that prioritises deployability. The research directions in Figure 11 translate the identified gaps into a structured agenda for the proper adoption and assimilation of HSCML.

5. Conclusion

This study addresses the critical gap in understanding how ML contributes to HSC through an LDA-guided, CIMO-structured, mixed-methods SLR. As the first systematic examination of data and model quality in HSCML, this review produced three novel deliverables:

1 a CIMO-structured characterisation of HSCML across contexts and six application domains;

Figure 11 Research agenda for HSCML based on review results



Source: Authors' own work

2 aggregated frameworks for data quality dimensions, model quality dimensions and factors influencing model quality; and
3 a deployability-focused, data-centric research agenda comprising 14 knowledge gaps and recommended directions across context, data and model dimensions.

Extending previous reviews, this study reveals the methodological impact mechanisms driving the disconnect between technical development and operational requirements, which are gaps that were overlooked by prior work.

This review offers structured guidance for multiple stakeholders. For *researchers*, the quality frameworks and research agenda provide systematic entry points for future work, particularly in the areas of uncertainty quantification, fairness-integrated and multi-model HSCML systems and deployment-oriented testing regimes identified across all six application domains. The context-focused review also identified underserved areas, such as LDCs, neglected disaster types and management phases and critical HSC stages and flows, such as procurement and finance.

For *practitioners* and *decision-makers,* HSCML operates in high-stakes environments where errors and biases translate directly into humanitarian harm. The systematic underinvestment in data quality, fairness, uncertainty quantification and explainability documented is not merely a technical deficiency but a governance risk that calls for institutional action, such as HSCML accountability frameworks, bias monitoring and demonstrable quality assurance as prerequisites for deployment, in line with obligations under international humanitarian law. The three synthesised frameworks from this literature review provide practical guidance for developing, evaluating and adopting HSCML with appropriate considerations.

Three limitations of this review should be acknowledged. Firstly, given the breadth of HSC domains identified, certain aspects, including ethical considerations, socio-political dimensions and organisational implementation factors, warrant dedicated systematic reviews beyond our scope. Secondly, this synthesis of published research would benefit from triangulation with primary data from HSCML practitioners, providing deeper insights into implementation challenges and operational barriers that the academic literature may not fully capture. Thirdly, given the field-level positioning of this review, specific pipeline design discussions and best practice recommendations for individual HSCML application domains were not covered. A dedicated sub-field SLR that examines pipeline-level practices in depth will be valuable, for which the established frameworks in this paper are intended to inform and scaffold.

Acknowledgements

The authors would like to thank Ms Nguyet Huynh from RMIT Vietnam Library for her valuable support in the curation of literature for this review and the editor-in-chief and two anonymous reviewers for their constructive comments and helpful suggestions on a previous draft of this paper.

References

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., Makarenkov, V. and Nahavandi, S. (2021), "A review of uncertainty quantification in deep learning: Techniques, applications and challenges", *Information Fusion*, Vol. 76, pp. 243-297, doi: 10.1016/j.inffus.2021.05.008.

Ahani, N., Andersson, T., Martinello, A., Teytelboym, A. and Trapp, A.C. (2021), "Placement optimization in refugee resettlement", *Operations Research*, Vol. 69 No. 5, pp. 1468-1486, doi: 10.1287/opre.2020.2093.

Ahmad, M., Tayyab, M. and Habib, M.S. (2025), "An enhanced deep reinforcement learning approach for efficient, effective, and equitable disaster relief distribution", *Engineering Applications of Artificial Intelligence*, Vol. 143, doi: 10.1016/j.engappai.2025.110002.

Alam, F., Ofli, F. and Imran, M. (2019), "Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of hurricanes Harvey, Irma, and maria", *Behaviour & Information Technology*, Vol. 39 No. 3, pp. 288-318, doi: 10.1080/0144929x.2019.1610908.

Almalki, F.A. and Angelides, M.C. (2019), "Deployment of an aerial platform system for rapid restoration of communications links after a disaster: a machine learning approach", *Computing*, Vol. 102 No. 4, pp. 829-864, doi: 10.1007/s00607-019-00764-x.

Altay, N., Heaslip, G., Kovacs, G., Spens, K., Tatham, P. and Vaillancourt, A. (2023), "Innovation in humanitarian logistics and supply chain management: a systematic review", *Annals of Operations Research*, Vol. 335 No. 3, pp. 1-23, doi: 10.1007/s10479-023-05208-6.

Amin Amani, M., Asumadu Sarkodie, S., Sheu, J.-B., Mahdi Nasiri, M. and Tavakkoli-Moghaddam, R. (2025), "A data-driven hybrid scenario-based robust optimization method for relief logistics network design", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 194, doi: 10.1016/j.tre.2024.103931.

Anjomshoae, A., Banomyong, R., Mohammed, F. and Kunz, N. (2022), "A systematic review of humanitarian supply chains performance measurement literature from 2007 to 2021", *International Journal of Disaster Risk Reduction*, Vol. 72, doi: 10.1016/j.ijdrr.2022.102852.

Araya, D., Podgorski, J. and Berg, M. (2023), "Groundwater salinity in the horn of Africa: spatial prediction modeling and estimated people at risk", *Environment International*, Vol. 176, p. 107925, doi: 10.1016/j.envint.2023.107925.

Barredo Arrieta, A., Díaz-Rodríguez, N., DEL Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F. (2020), "Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI", *Information Fusion*, Vol. 58, pp. 82-115, doi: 10.1016/j.inffus.2019.12.012.

Behl, A., Bhardwaj, S., Jayawardena, N., Pereira, V. and Roohanifar, M. (2026), "Grass is always dark(er) on the other side: exploring the dark side of artificial intelligence humanitarian supply chain operations", *Technological Forecasting and Social Change*, Vol. 224, doi: 10.1016/j.techfore.2025.124484.

Bhullar, G., Khullar, A., Kumar, A., Sharma, A., Pannu, H.S. and Malhi, A. (2022), "Time series sentiment analysis (SA) of relief operations using social media (SM) platform for efficient resource management", *International Journal of Disaster Risk Reduction*, Vol. 75, doi: 10.1016/j.ijdrr.2022.102979.

Biswas, S., Kumar, D., Hajiaghaei-Keshteli, M. and Bera, U. K. (2024), "An AI-based framework for earthquake relief demand forecasting: a case study in Türkiye", *International Journal of Disaster Risk Reduction*, Vol. 102, doi: 10.1016/j.ijdrr.2024.104287.

Caton, S. and Haas, C. (2024), "Fairness in machine learning: a survey", *ACM Computing Surveys*, Vol. 56 No. 7, p. 166, doi: 10.1145/3616865.

Chen, J. and Cho, Y.K. (2022), "CrackEmbed: point feature embedding for crack segmentation from disaster site point clouds with anomaly detection", *Advanced Engineering Informatics*, Vol. 52, doi: 10.1016/j.aei.2022.101550.

Chi, G., Fang, H., Chatterjee, S. and Blumenstock, J.E. (2022), "Microestimates of wealth for all low- and Middle-income countries", *Proceedings of the National Academy of Sciences*, Vol. 119 No. 3, doi: 10.1073/pnas.2113658119.

Chiappetta Jabbour, C.J., Sobreiro, V.A., Lopes DE Sousa Jabbour, A.B., DE Souza Campos, L.M., Mariano, E.B. and Renwick, D.W.S. (2017), "An analysis of the literature on humanitarian logistics and supply chain management: paving the way for future studies", *Annals of Operations Research*, Vol. 283 Nos 1-2, pp. 289-307, doi: 10.1007/s10479-017-2536-x.

Chuang, J., Manning, C.D. and Heer, J. (2012), "Termite: visualization techniques for assessing textual topic models", *Proceedings of the International Working Conference on Advanced Visual Interfaces. Capri Island, Association for Computing Machinery, Italy*.

DE Camargo Fiorini, P., Chiappetta Jabbour, C.J., Lopes DE Sousa Jabbour, A.B. and Ramsden, G. (2022), "The human side of humanitarian supply chains: a research agenda and systematization framework", *Annals of Operations Research*, Vol. 319 No. 1, pp. 911-936, doi: 10.1007/s10479-021-03970-z.

DE Santi, M., Khan, U.T., Arnold, M., Fesselet, J.-F. and Ali, S.I. (2021), "Forecasting point-of-consumption chlorine residual in refugee settlements using ensembles of artificial neural networks", *Npj Clean Water*, Vol. 4 No. 1, doi: 10.1038/s41545-021-00125-2.

Denyer, D. and Tranfield, D. (2009), "Producing a systematic review", *The Sage Handbook of Organizational Research Methods*, Sage Publications Ltd., Thousand Oaks, ISBN: 978-1-4129-3118-2.

Diez-Tomillo, J., Martinez-Alpiste, I., Golcarenarenji, G., Wang, Q. and Alcaraz-Calero, J.M. (2024), "Efficient CNN-based low-resolution facial detection from UAVs", *Neural Computing and Applications*, Vol. 36 No. 11, pp. 5847-5860, doi: 10.1007/s00521-023-09401-3.

Dorsey, J. (2026), "The erosion of human(e) judgement in targeting? Quantification logics, AI-enabled decision support systems and proportionality assessments in IHL", *International Review of the Red Cross*, Vol. 107 No. 930, pp. 1-31, doi: 10.1017/s1816383125100969.

Fan, C., Wu, F. and Mostafavi, A. (2020), "A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters", *IEEE Access*, Vol. 8, pp. 10478-10490, doi: 10.1109/access.2020.2965550.

Faraci, G., Rizzo, S.A. and Schembra, G. (2023), "Green edge intelligence for smart management of a FANET in Disaster-Recovery scenarios", *IEEE Transactions on Vehicular Technology*, Vol. 72 No. 3, pp. 3819-3831, doi: 10.1109/tvt.2022.3217331.

Feng, Y., Wang, X., Wang, D., Yin, Y. and Ignatius, J. (2025), "An interpretable two-stage adaptive deep learning model for humanitarian aid information prediction and emergency response support", *Technological Forecasting and Social Change*, Vol. 219, p. 124293, doi: 10.1016/j.techfore.2025.124293.

Fuqua, D. and Hespeler, S. (2022), "Commodity demand forecasting using modulated rank reduction for humanitarian logistics planning", *Expert Systems with Applications*, Vol. 206, doi: 10.1016/j.eswa.2022.117753.

Galera-Zarco, C. and Floros, G. (2023), "A deep learning approach to improve built asset operations and disaster management in critical events: an integrative simulation model for quicker decision making", *Annals of Operations Research*, Vol. 339 Nos 1-2, pp. 573-612, doi: 10.1007/s10479-023-05247-z.

Garcia, C., Rabadi, G., Abujaber, D. and Seck, M. (2023), "Supporting humanitarian crisis decision making with reliable intelligence derived from social media using AI", *Journal of Homeland Security and Emergency Management*, Vol. 20 No. 2, pp. 97-131, doi: 10.1515/jhsem-2021-0042.

Ghannad, P., Lee, Y.-C. and Choi, J.O. (2021), "Prioritizing postdisaster recovery of transportation infrastructure systems using multiagent reinforcement learning", *Journal of Management in Engineering*, Vol. 37 No. 1, doi: 10.1061/(asce)me.1943-5479.0000868.

Gong, Y., LI, X., Belabbes, S., Dell'oro, L. and Ru, Y. (2024), "Multidimensional geographic factors behind conflicts: a case study in Sudan", *International Journal of Digital Earth*, Vol. 17 No. 1, doi: 10.1080/17538947.2024.2426524.

Gour, A., Aggarwal, S. and Kumar, S. (2022), "Lending ears to unheard voices: an empirical analysis of user-generated content on social media", *Production and Operations Management*, Vol. 31 No. 6, pp. 2457-2476, doi: 10.1111/poms.13732.

Grant, M.J. and Booth, A. (2009), "A typology of reviews: an analysis of 14 review types and associated methodologies", *Health Information & Libraries Journal*, Vol. 26 No. 2, pp. 91-108, doi: 10.1111/j.1471-1842.2009.00848.x.

Harden, A. and Thomas, J. (2010), "Mixed methods and systematic reviews: Examples and emerging issues", *in* Tashakkori, A. and Teddlie, C. (Eds), *SAGE Handbook of Mixed Methods in Social & Behavioral Research*, SAGE Publications, Inc., Thousand Oaks, ISBN: 9781412972666.

Hazarika, B., Singh, P., Singh, K., Cotton, S.L., Shin, H., Dobre, O.A. and Duong, T.Q. (2025), "Generative AI-Augmented graph reinforcement learning for adaptive UAV swarm optimization", *IEEE Internet of Things Journal*, Vol. 12 No. 8, doi: 10.1109/jiot.2025.3529904.

Heuschmid, D., Wacker, O., Zimmermann, Y., Penava, P. and Buettner, R. (2025), "Advancements in landmine detection: deep Learning-Based analysis with thermal drones", *IEEE Access*, Vol. 13, pp. 91777-91794, doi: 10.1109/access.2025.3572196.

Hssini, I. (2025), "Distributionally robust and Learning-Based vehicle routing for disaster blood logistics", *IEEE Access*, Vol. 13, pp. 119538-119550, doi: 10.1109/access.2025.3584213.

Hu, D., Chen, J. and Li, S. (2022), "Reconstructing unseen spaces in collapsed structures for search and rescue via deep learning based radargram inversion", *Automation in Construction*, Vol. 140, doi: 10.1016/j.autcon.2022.104380.

Hu, S., Dong, Z.S. and Dai, R. (2024), "A machine learning based sample average approximation for supplier selection with option contract in humanitarian relief", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 186, doi: 10.1016/j.tre.2024.103531.

Hu, D., Li, S., Du, J. and Cai, J. (2023), "Automating building damage reconnaissance to optimize drone mission planning for disaster response", *Journal of Computing in Civil Engineering*, Vol. 37 No. 3, doi: 10.1061/(asce)cp.1943-5487.0001061.

Jana, D., Malama, S., Narasimhan, S. and Taciroglu, E. (2023), "Edge-based graph neural network for ranking critical road segments in a network", *Plos One*, Vol. 18 No. 12, p. e0296045, doi: 10.1371/journal.pone.0296045.

Kembro, J., Kunz, N., Frennesson, L. and Vega, D. (2024), "Revisiting the definition of humanitarian logistics", *Journal of Business Logistics*, Vol. 45 No. 2, doi: 10.1111/jbl.12376.

Khatoon, S., A. Alshamari, M., Asif, A., Maruf Hasan, M., Abdou, S., Mostafa Elsayed, K. and Rashwan, M. (2021), "Development of social media analytics system for emergency event detection and crisis management", *Computers, Materials & Continua*, Vol. 68 No. 3, pp. 3079-3100, doi: 10.32604/cmc.2021.017371.

Khosravi, R. and Rabbani, M. (2026), "A telemedicine-oriented framework for scenario-based disaster management: integrating machine learning and multi-criteria decision making", *Applied Soft Computing*, Vol. 186, p. 114274, doi: 10.1016/j.asoc.2025.114274.

Kiavash, P., Tanaltay, A. and Tabatabaei, R.A. (2026), "Can social media predict demand in humanitarian crises? A case study of the 2023 Türkiye earthquake", *Technology in Society*, Vol. 84, p. 103054, doi: 10.1016/j.techsoc.2025.103054.

Kondraganti, A., Narayanamurthy, G. and Sharifi, H. (2022), "A systematic literature review on the use of big data analytics in humanitarian and disaster operations", *Annals of Operations Research*, Vol. 335 No. 3, pp. 1-38, doi: 10.1007/s10479-022-04904-z.

Kumar, A. and Singh, J.P. (2019), "Location reference identification from tweets during emergencies: a deep learning approach", *International Journal of Disaster Risk Reduction*, Vol. 33, pp. 365-375, doi: 10.1016/j.ijdrr.2018.10.021.

Kumar, P., Singh, R.K. and Shahgholian, A. (2022), "Learnings from COVID-19 for managing humanitarian supply chains: systematic literature review and future research directions", *Annals of Operations Research*, Vol. 335 No. 3, pp. 1-37, doi: 10.1007/s10479-022-04753-w.

Lee, H.-R. and Lee, T. (2021), "Multi-agent reinforcement learning algorithm to solve a partially-observable multi-agent problem in disaster response", *European Journal of Operational Research*, Vol. 291 No. 1, pp. 296-308, doi: 10.1016/j.ejor.2020.09.018.

Lin, A., WU, H., Liang, G., Cardenas-Tristan, A., Wu, X., Zhao, C. and Li, D. (2020), "A big data-driven dynamic estimation model of relief supplies demand in urban flood disaster", *International Journal of Disaster Risk Reduction*, Vol. 49, doi: 10.1016/j.ijdrr.2020.101682.

Maric, J., Galera-Zarco, C. and Opazo-Basaez, M. (2022), "The emergent role of digital technologies in the context of humanitarian supply chains: a systematic literature review", *Annals of Operations Research*, Vol. 319 No. 1, pp. 1003-1044, doi: 10.1007/s10479-021-04079-z.

Nabavi, S.M., Vahdani, B., Nadjafi, B.A. and Adibi, M.A. (2022), "Synchronizing victim evacuation and debris removal: a data-driven robust prediction approach", *European Journal of Operational Research*, Vol. 300 No. 2, pp. 689-712, doi: 10.1016/j.ejor.2021.09.051.

Nadi, A. and Edrisi, A. (2017), "Adaptive multi-agent relief assessment and emergency response", *International Journal of Disaster Risk Reduction*, Vol. 24, pp. 12-23, doi: 10.1016/j.ijdrr.2017.05.010.

Nawazish, M., Nandakumar, M.K. and Mateen, A. (2023), "Are humanitarian supply chains sustainable? A systematic review and future research themes", *Benchmarking: An International Journal*, Vol. 31 No. 8, pp. 2565-2601, doi: 10.1108/bij-01-2023-0036.

Nguyen, S., Fu, X., Ogawa, D. and Zheng, Q. (2023), "An application-oriented testing regime and multi-ship predictive modeling for vessel fuel consumption prediction", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 177, doi: 10.1016/j.tre.2023.103261.

Nguyen, L., Yang, Z., Li, J., Pan, Z., Cao, G. and Jin, F. (2022), "Forecasting people's needs in hurricane events from social network", *IEEE Transactions on Big Data*, Vol. 8 No. 1, pp. 229-240, doi: 10.1109/tbdata.2019.2941887.

Nguyen, S., Fu, X., Zhao, L., Xu, H., Zhang, X., Li, N., Zhang, W., Yin, X.F., Daichi, O., Koh, J. and Zheng, Q. (2026), "AI-driven Just-In-Time coordination between port and ships: advancing Maritime decarbonization", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 207, doi: 10.1016/j.tre.2025.104569.

OCHA (2024), *Briefing Note on Artificial Intelligence and the Humanitarian Sector*, United Nations Office for the Coordination of Humanitarian Affairs, UN.

Pan, J.-S., Song, P.-C., Chu, S.-C., Snášel, V. and Watada, J. (2025), "Machine Learning-Enabled evolutionary Two-Stage stochastic programming", *IEEE Transactions on Emerging Topics in Computational Intelligence*, pp. 1-12, doi: 10.1109/tetci.2025.3532539.

Peng, W., Wang, D., Yin, Y. and Cheng, T.C.E. (2025), "Multi-agent deep reinforcement learning-based truck-drone collaborative routing with dynamic emergency response", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 195, doi: 10.1016/j.tre.2025.103974.

Polushko, V., Jenal, A., Bongartz, J., Weber, I., Hatic, D., Rösch, R., März, T., Rauhut, M. and Weinmann, A. (2024), "BlessemFlood21: a High-Resolution georeferenced dataset for advanced analysis of river flood scenarios", *IEEE Access*, Vol. 12, pp. 176389-176405, doi: 10.1109/access.2024.3487413.

Quiliche, R., Santiago, B., Baião, F.A. and Leiras, A. (2023), "A predictive assessment of households' risk against disasters caused by cold waves using machine learning", *International Journal of Disaster Risk Reduction*, Vol. 98, doi: 10.1016/j.ijdrr.2023.104109.

Rahman, M., Chen, N., Islam, M.M., Mahmud, G.I., Pourghasemi, H.R., Alam, M., Rahim, M.A., Baig, M.A., Bhattacharjee, A. and Dewan, A. (2021), "Development of flood hazard map and emergency relief operation system using hydrodynamic modeling and machine learning algorithm", *Journal of Cleaner Production*, Vol. 311, doi: 10.1016/j.jclepro.2021.127594.

Saleem, S., Hasan, N., Khattar, A., Jain, P.R., Gupta, T.K. and Mehrotra, M. (2024), "DeLTran15: a deep lightweight Transformer-Based framework for multiclass classification of disaster posts on X", *IEEE Access*, Vol. 12, pp. 153676-153693, doi: 10.1109/access.2024.3478790.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F. and Dennison, D. (2015), "Hidden technical debt in machine learning systems", *Proceedings of the 29th International Conference on Neural Information Processing Systems – Volume 2. Montreal, Canada*: *MIT Press*.

Sentia, P.D., Abdul Shukor, S., Wahab, A.N.A. and Mukhtar, M. (2023), "Logistic distribution in humanitarian supply chain management: a thematic literature review and future research", *Annals of Operations Research*, Vol. 323 Nos 1-2, pp. 175-201, doi: 10.1007/s10479-023-05232-6.

Sievert, C. and Shirley, K. (2014), *LDAvis: A Method for Visualizing and Interpreting Topics*, Association for Computational Linguistics, Baltimore, MD, pp. 63-70.

Song, X., Shibasaki, R., Yuan, N.J., Xie, X., Li, T. and Adachi, R., (2017), "DeepMob: learning deep knowledge of human emergency behavior and mobility from big and heterogeneous data", *ACM Transactions on Information Systems*, Vol. 35 No. 4, pp. 1-19, doi: 10.1145/3057280.

Speth, S., Gonçalves, A., Rigault, B., Suzuki, S., Bouazizi, M., Matsuo, Y. and Prendinger, H. (2022), "Deep learning with RGB and thermal images onboard a drone for monitoring operations", *Journal of Field Robotics*, Vol. 39 No. 6, pp. 840-868, doi: 10.1002/rob.22082.

Taouktsis, X. and Zikopoulos, C. (2024), "A decision-making tool for the determination of the distribution center location in a humanitarian logistics network", *Expert Systems with Applications*, Vol. 238, doi: 10.1016/j.eswa.2023.122010.

Terti, G., Ruin, I., Gourley, J.J., Kirstetter, P., Flamig, Z., Blanchet, J., Arthur, A. and Anquetin, S. (2019), "Toward probabilistic prediction of flash flood human impacts", *Risk Analysis*, Vol. 39 No. 1, pp. 140-161, doi: 10.1111/risa.12921.

Ülkü, M.A., Bookbinder, J.H. and Yun, N.Y. (2024), "Leveraging industry 4.0 technologies for sustainable humanitarian supply chains: evidence from the extant literature", *Sustainability*, Vol. 16 No. 3, doi: 10.3390/su16031321.

Ullah, I., Khan, S., Imran, M. and Lee, Y.-K. (2021), "RweetMiner: automatic identification and categorization of help requests on twitter during disasters", *Expert Systems with Applications*, Vol. 176, doi: 10.1016/j.eswa.2021.114787.

VAN Steenbergen, R., Mes, M. and VAN Heeswijk, W. (2023), "Reinforcement learning for humanitarian relief distribution with trucks and UAVs under travel time uncertainty", *Transportation Research Part C: Emerging Technologies*, Vol. 157, doi: 10.1016/j.trc.2023.104401.

Vanvuchelen, N., DE Boeck, K. and Boute, R.N. (2024), "Cluster-based lateral transshipments for the Zambian health supply chain", *European Journal of Operational Research*, Vol. 313 No. 1, pp. 373-386, doi: 10.1016/j.ejor.2023.08.005.

Vlachos, I. and Reddy, P.G. (2025), "Machine learning in supply chain management: systematic literature review and future research agenda", *International Journal of Production Research*, Vol. 63 No. 16, pp. 1-30, doi: 10.1080/00207543.2025.2466062.

Wen, B.-J. and Chen, Y.-H. (2023), "Night-Time measurement and skeleton recognition using unmanned aerial vehicles equipped with LiDAR sensors based on Deep-Learning algorithms", *IEEE Sensors Journal*, Vol. 23 No. 19, pp. 23474-23485, doi: 10.1109/jsen.2023.3302524.

Wu, H., Ghadami, A., Bayrak, A.E., Smereka, J.M. and Epureanu, B.I. (2021), "Impact of heterogeneity and risk aversion on task allocation in Multi-Agent teams", *IEEE Robotics and Automation Letters*, Vol. 6 No. 4, pp. 7065-7072, doi: 10.1109/lra.2021.3097259.

Xia, H., Wu, J., Yao, J., Zhu, H., Gong, A., Yang, J., Hu, L. and Mo, F. (2023), "A deep learning application for building damage assessment using Ultra-High-Resolution remote sensing imagery in Turkey earthquake", *International Journal of Disaster Risk Science*, Vol. 14 No. 6, pp. 947-962, doi: 10.1007/s13753-023-00526-6.

Xiong, R., Peng, H., Chen, X. and Shuai, C. (2024), "Machine learning-enhanced evaluation of food security across 169 economies", *Environment, Development and Sustainability*, Vol. 26 No. 10, pp. 26971-27000, doi: 10.1007/s10668-024-05212-1.

YAĞCıOĞLU, M. (2025), "Machine learning based dynamic resource sharing and frequency reuse in 5G hetnets with dronecells", *Computer Networks*, Vol. 258, doi: 10.1016/j.comnet.2025.111046.

Yu, L., Zhang, C., Jiang, J., Yang, H. and Shang, H. (2021), "Reinforcement learning approach for resource allocation in humanitarian logistics", *Expert Systems with Applications*, Vol. 173, doi: 10.1016/j.eswa.2021.114663.

Zhang, Z., Ma, Y. and Liu, P. (2024b), "A global multimodal flood event dataset with heterogeneous text and multi-source remote sensing images", *Big Earth Data*, Vol. 9 No. 3, pp. 1-27, doi: 10.1080/20964471.2024.2358615.

Zhang, T., Lauras, M., Zacharewicz, G., Rabah, S. and Benaben, F. (2024a), "Coupling simulation and machine learning for predictive analytics in supply chain management", *International Journal of Production Research*, Vol. 62 No. 23, pp. 8397-8414, doi: 10.1080/00207543.2024.2342019.

Zhang, W., Wen, J., Dong, H., Han, Q. and Du, X. (2025), "Post-earthquake functionality and resilience prediction of bridge networks based on data-driven machine learning method", *Soil Dynamics and Earthquake Engineering*, Vol. 190, doi: 10.1016/j.soildyn.2024.109127.

Zhao, Z., Zhou, X., Zheng, Y., Meng, T. and Fang, D. (2024b), "Enhancing infrastructural dynamic responses to critical residents' needs for urban resilience through machine learning and hypernetwork analysis", *Sustainable Cities and Society*, Vol. 106, doi: 10.1016/j.scs.2024.105366.

Zhao, X., Wang, H., Morikawa, S., Wang, S., Tang, J., Sugiyama, S. and Sriwarnasinghe, S.M. (2025), "Mapping non-profits participation in disaster mitigation: a data-driven study of functional diversity, spatial patterns and driving factors of China's emergency management nonprofit organizations", *International Journal of Disaster Risk Reduction*, Vol. 118, doi: 10.1016/j.ijdrr.2025.105252.

Zhao, Q., Tan, X., Meng, Q., Gao, L., Zhang, L., Lei, X., Shi, W. and Zhu, M. (2024a), "Satellite and AI monitoring humanitarian crises in the Gaza Strip during the early stage of Israeli–Palestinian conflict", *International Journal of Digital Earth*, Vol. 17 No. 1, doi: 10.1080/17538947.2024.2430678.

Zheng, Y.-J., Yu, S.-L., Song, Q., Huang, Y.-J., Sheng, W.-G. and Chen, S.-Y. (2022), "Co-Evolutionary fuzzy deep transfer learning for disaster relief demand forecasting", *IEEE Transactions on Emerging Topics in Computing*, Vol. 10 No. 3, pp. 1361-1373, doi: 10.1109/tetc.2021.3085337.

Zhu, Y. and Wang, S. (2023), "Cooperative aerial and ground vehicle trajectory planning for post-disaster communications: an attention-based learning approach", *Vehicular Communications*, Vol. 42, doi: 10.1016/j.vehcom.2023.100631.

Corresponding author

Son Nguyen can be contacted at: son.nguyen74@rmit.edu.vn