Progress in Disaster Science 25 (2025) 100404

Contents lists available at ScienceDirect

Progress in Disaster Science

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

Improving operational use of post-disaster damage assessment for Urban Search and Rescue by integrated graph-based multimodal remote sensing data analysis

Sivasakthy Selvakumaran [a,b,*,1], Iain Rolland [a], Luke Cullen [a], Rob Davis [b,c], Joshua Macabuag [b], Charbel Abou Chakra [d], Nanor Karageozian [d], Amir Gilani [e], Christian Geiß [f,g], Miguel Bravo-Haro [h], Andrea Marinoni [i]

[a] Engineering Department, University of Cambridge, United Kingdom
[b] Search and Rescue Assistance in Disasters (SARAID), United Kingdom
[c] UCL Institute for Risk and Disaster Reduction, University College London, United Kingdom
[d] Urban Analysis and Policy Unit, United Nations Human Settlements Programme (UN-Habitat), Lebanon
[e] Miyamoto International, USA
[f] German Remote Sensing Data Center, German Aerospace Center (DLR), Germany
[g] Department of Geography, University of Bonn, Bonn 53115, Germany
[h] City St George's, University of London, United Kingdom
[i] Department of Physics and Technology, UiT the Arctic University of Norway, Norway

ARTICLE INFO

Keywords:
Disaster management
Post-disaster
Urban Search and Rescue (USAR)
Remote sensing
Graph-based data analysis
Machine learning

ABSTRACT

This work investigates the application of remote sensing technologies within the specific operational context of emergency urban search and rescue (USAR) efforts post-disaster. We thoroughly investigate two innovative methodologies, each tailored to meet distinct operational goals in a USAR setting. The first employs a belief propagation framework that is designed to provide prompt and robust initial damage assessments using sparse data, with the capability to incorporate additional on-site information as it becomes available. The second methodology introduces a modified graph convolutional network to quantify the uncertainty levels inherent in damage classification tasks. Three case studies were considered, using damage assessment data from the 2020 Beirut explosion, the 2021 Haiti earthquake and the 2023 Türkiye-Syria earthquake. Our experimental results demonstrate the potential of these approaches to achieve operational objectives, particularly in terms of robustness and scalability in disaster scenarios. The versatility offered by graph-based methodologies establishes a solid foundation for addressing these dynamic challenges, suggesting a promising direction for continued research in this field.

1. Introduction

Disasters triggered by natural hazards or technological accidents create a threat for human safety and the built environment. Rapid and accurate building and infrastructure damage assessment is critical for humanitarian response. The amount of damage assessment information available, as well as the specific purpose for needing that information, changes over time as the response evolves [1]. In the immediate aftermath of a disaster, there is little information on the ground, with further complications arising if communication and electricity networks are down, or physical access to the disaster zone is limited. In immediate search and rescue activities, there is a need for a rapid assessment in order to plan the deployment of further resources. In later stages, as information becomes more available, it can be used for planning specific recovery activities. Collecting data on the ground can be difficult, dangerous and time-consuming. In this context, satellite remote sensing can provide additional information through wide area observations and analysis of imagery collected remotely [2,3,4].

There are many different forms of remote sensing data, with temporal and spatial resolutions continually improving over time. With this,

* Corresponding author at: Engineering Department, University of Cambridge, United Kingdom.
E-mail address: ss683@cam.ac.uk (S. Selvakumaran).
[1] Postal Address: Cambridge University Engineering Department, Trumpington Street, Cambridge, CB2 1PZ.

https://doi.org/10.1016/j.pdisas.2025.100404
Received 10 April 2024; Received in revised form 30 December 2024; Accepted 1 January 2025
Available online 3 January 2025

various methods for using remote sensing for building damage assessment have been designed [5]. Optical data has long been used for pure visual interpretation, manually comparing images before and after events [6]. The recent technological advancements in optical resolution have enabled relative success in deploying pattern recognition techniques based on convolutional neural network models which make use of very high resolution (VHR) data (1 m by 1 m pixel resolution or often much higher) for building segmentation and damage recognition [7]. Other forms of remote sensing data on the electromagnetic spectrum have also been investigated. Synthetic Aperture Radar (SAR) is an active imaging format whose properties of intensity and interferometric coherence can be used in various post-event damage assessment applications [8,9]. It is particularly useful, as active imaging allows the satellite to be used day and night, and it is not affected by cloud cover. Thermal satellite data has been used in operational systems to detect hotspots caused by wildfires and burn scars following events [10]. Multispectral and hyperspectral data sources have been used in a wide range of automated segmentation and classification studies [11].

Multi-sensor methods have also been applied in tasks related to damage classification. Multimodal remote sensing means using different sensors, possibly operating at different frequencies and/or spatial resolutions and with different time intervals between acquisitions over a given region [12,2]. Having multiple and heterogeneous image sources that are available over the same geographical region improve classification by allowing sources to be combined/fused. The benefit of using multimodal data is the ability to exploit the different aspects captured by different sources. For example, optical data provides a view similar to looking at an overhead view of a city using one's own eyes, whereas SAR imagery tells us something different by measuring how radar is reflected and collected by a satellite in a side-looking viewing geometry (i.e. potentially collecting more information about the side facades of buildings compared to nadir optical images). To exploit the benefits of multiple modes, different approaches have been taken. Some analyse each individual mode and then combine the different results, while others aim at data fusion earlier in the analysis, such that the different datasets can interact and inform each other [13]. The progress made in satellite missions means that we are able to access increasingly improving resolutions and quality of data. On the other hand, methods are required to exploit the data quickly and effectively, to make sure that relevant information reaches decision-makers in a timely manner.

Various approaches have been used with regard to multimodal remote sensing for post-disaster damage assessment. For example, Gomez-Chova et al. [14] use SAR image pairs together with optical images (four spectral channels), acquired over an area of Port-au-Prince, Haiti, before and after the 12 January 2010 earthquake and employ an unsupervised approach by using change vector analysis of the before and after images. Meanwhile, other studies, such as Brunner et al. [15], make use of a supervised learning approach to exploit the different properties of VHR SAR and VHR optical data collected over Yingxiu, China, which was heavily damaged in the Sichuan earthquake of 12 May 2008. They use pre-event VHR optical imagery to estimate the 3-D parameters of a building and use this with the acquisition parameters of the VHR SAR scene to predict the expected signature of a building without damage, before finally comparing this to the actual post-event SAR image. Another example by Geiß et al. [16] is the use of various multispectral imaging formats together with digital elevation models to estimate seismic building structural types in Padang, Indonesia, prior to a potential earthquake disaster through the use of feature selection, outlier detection and a supervised classification method using synthetic samples generated as part of the study. More recently, the feasibility of deep learning strategies for assessing damage has been investigated (refer to, for example, Adriano et al. [17]), especially considering the various combinations of pre- and post-disaster remote sensing records (mainly optical and SAR). In particular, it is shown that exploiting the diversity of the heterogeneous remote sensing data leads to improved information extraction on the region affected by disasters.

Remote sensing is a valuable tool in post-disaster operations. However, the achievements that have been made in the research community are not necessarily striving to step side-by-side with the needs of those working on the ground over the progression of an immediate disaster response. For example, the International Search and Rescue Advisory Group (INSARAG), a global network under the United Nations umbrella, has established the INSARAG Coordination and Management System (ICMS). This system is designed for the real-time collection and visualization of operational data from Urban Search and Rescue (USAR) teams during disaster response. Its primary aim is to provide a comprehensive and timely situational awareness that is critical for informed operational decision-making. Integrating remote sensing data into the ICMS could potentially enhance this situational awareness, offering an additional layer of information to inform critical decisions. However, despite the potential benefits, there currently are no remote-sensing tools that meet the necessary criteria of accuracy, robustness, and speed to reliably contribute to the ICMS for real-time operational decision support in disaster scenarios. In this regard, there is still work to be done in understanding what is needed along the progression of a disaster scenario and work to be done in developing methods that support these evolving needs.

In this work, we introduce an architecture to enhance the information provided to USAR operational organizations when responding to a disaster event. In this respect, we consider that a key factor for the success of USAR operations is to provide timely information to ensure an effective and efficient response of USAR units. As such, we developed a two-stage system that is able to support the decision-making processes of units *in the field*, also when scarce computational power is available and fast decisions have to be taken. The main two steps of this architecture are:

- A scheme based on belief propagation approach, aiming to provide a fast and reliable classification of damaged and undamaged buildings by analyzing pre-event remote sensing data and the in-situ observations that would be provided by USAR units in the field;
- An uncertainty-aware graph-based data analysis scheme which is able to quantify the degree of confidence that is possible to associate with each analyzed building, so to support the implementation of better-informed decision making systems directly at the event site.

These algorithms are implemented in order to guarantee that the data-driven damage assessment could be conducted in the field, hence ensuring light power consumption, fast response, and high interpretability in order to support their use by USAR units (who typically show limited experience in data analysis and processing platforms). The proposed architecture is able to provide robust estimates of damage extent and severity also on platforms with limited computational power, hence facilitating and supporting the better-informed planning by USAR organizations on the field, which would lead ultimately to efficient unit deployments on ground and more prompt interventions to support the local communities in hard times.

The remainder of this paper is organized as follows: Section 2 presents some of the operational considerations of the end users of post-disaster damage assessment information and decision-makers in response efforts, which provide context required for developing effective remote sensing methodologies in such scenarios. Section 3 gives information about three specific case studies in Lebanon, Haiti and Turkey from which in-situ information has been collected and on which the methodologies are assessed. Section 4 outlines the methods employed, and Sections 5 and 6 discuss the results and provide conclusions from this work.

2. Post-disaster response stages and requirements

Following a sudden-onset disaster causing large-scale structural collapse, there is immediate anticipation about the potential

development of the situation. Decisions on how to act must be made by individuals and organizations – in the public, non-profit, private, etc. realms – involved in response operations. Disaster management is required to organize and direct resources and responders, and it is a constantly evolving process [18]. Systematic risk identification and evaluation, implementation of countermeasures, awareness raising, and decision support systems should be employed before, as well as during and after a disaster to reduce the vulnerabilities of societies and the infrastructure that supports them [19]. Ideally, the local authority would have plans in place in the event of a disaster, with structures and measures in place and the local population prepared. Even so, an immediate assessment of the damages is required, from transport links to buildings – and remote sensing can be a supporting tool.

During a crisis resulting from an acute disaster like an earthquake or tsunami, while the situation in the disaster area may still be chaotic and unclear, rapid mapping activities using up-to-date satellite data can provide valuable information to support emergency response actions [20]. The country where the disaster has taken place designates a Local Emergency Management Agency (LEMA), which is often a branch of the local government or military, which coordinates the response. They are best placed to understand the existing situation prior to the disaster, and have critical information about the area, building mapping information and structural understanding. LEMA then co-ordinates emergency services and Urban Search and Rescue (USAR) teams (local, national and possibly international) to conduct search and rescue [21] as well as emergency and temporary structural works. The initial search and rescue phase varies, lasting from hours or even days after the disaster depending on circumstances. Rapid response is needed to save lives placed in imminent danger immediately following the event; within a few days the work shifts towards a different phase focused on providing support to the survivors.

An immediate prioritization of life-saving activities is required, including the direction of resources for USAR activities and structural damage assessment [22]. Sector prioritization is key to reaching the most affected areas quickly, potentially saving lives; hence, decisions must be justifiable and accountable in case of later inquiry (Fig. 1). Prioritization is based both upon prior knowledge of an area's vulnerability and initial damage assessments from ground reports or aerial photography [23]. Remote sensing provides an opportunity to aid this initial critical decision-making process by providing additional information about what is happening over a wide geographic area, before sites can be accessed (which can be more difficult in circumstances where transport networks have been compromised by the disaster, for example).

Rescue services can get a very rough visual overview of the current state of the affected area by using satellite imagery acquired immediately after the event. Comparison of images acquired before and after the event identify major changes on the ground caused by the disaster, to highlight regions that are likely to be most badly affected and identify passable routes for rescue and humanitarian workers. Satellite images can also help to identify areas that are suitable for setting up base camps and emergency shelters. A limitation of optical images is that the presence of clouds can obscure the image. Many damage proxy maps that are released soon after disaster events often make use of SAR data collected before and after the event, with some validation through comparing local media information and photos as they become available. Damage proxy maps are used as a guide to identify damaged areas; however, such maps are a rough guide and are quite coarse. Furthermore, they may be less reliable in the case of areas that have vegetation. Pixels marking damage over vegetated areas may be false positives, and conversely the lack of marked damage areas over vegetated areas is not clear indication that damage has not occurred.

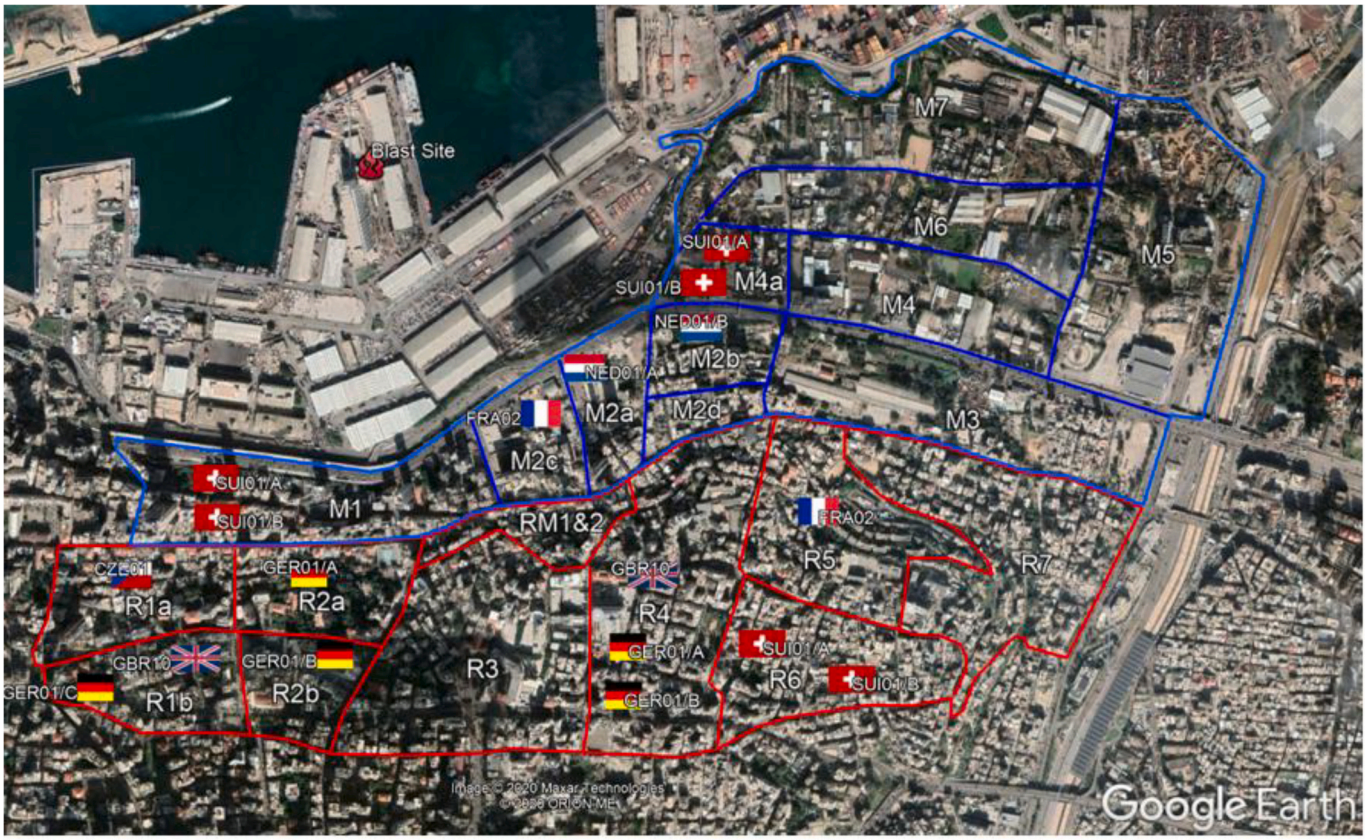

Fig. 1. Example of a sectorization map. This example shows how regions around the blast site in Beirut in 2020 were marked out with each region allocated for different international search and rescue teams working alongside local emergency teams. This is done in order to coordinate the search and rescue response. Map created by and reproduced with permission of SARAID.

The requirements of remote sensing information therefore need more than a simple estimate on whether there is damage or not. Disaster response teams offer abilities beyond searching for and rescuing victims, and key among these is effective damage assessment. This is necessary for several reasons [24]:

- To understand the nature and extent (scale and spread) of the disaster's effects.
- To manage risks associated with first and subsequent responses.
- To quantify needs for recovery from the disaster (e.g. resources and finances).
- To determine how soon reoccupation of buildings and restoration of the local infrastructure can be achieved.

The requirements for this information change depending on the stage of response, and the specific purpose of its use. For example, for remote loss estimates [25], insurance claims and determining repair works, this work needs to be accurate and detailed but is not needed within the first few hours of a response. Conversely, the initial hours of the response would require information as quickly as possible, and some information is better than an absence of any on-the-ground knowledge and can inform about access and supports decisions around sectorization. Therefore, the specific requirements for remote sensing data analysis change as priorities shift from USAR to damage assessment and beyond. This paper looks at potential methods for supporting the USAR scenario specifically, i.e. conditions with limited training material (limited information on the ground), changing amounts of information requiring continuously updated analysis (further data being collected on the ground and reported back) and the need for information immediately to make life-saving decisions (in minutes rather than hours or days). Thus, the methods considered must be robust, flexible and with minimal computational effort and time.

3. Study areas and data

In this study, we look at three different disaster scenarios covering disasters as a result of both natural hazards and man-made technological accidents. Section 3.1 presents the context and data used to study the port explosion in Beirut in 2020, and Sections 3.2 and 3.3 present the Haiti 2021 and Türkiye 2023 earthquakes, respectively. In each case the satellite data used for each study is presented, alongside the information available following the response about structural damage classifications and other additional information.

3.1. Beirut port explosion, 2020

On 4 August 2020, breaking news of a massive explosion in Beirut Port began to spread across the world. It was widely reported to have been caused by detonation of a large quantity of ammonium nitrate, following a fire in the warehouse where it was being stored. The explosion killed more than 200, injured over 6000, and displaced around 300,000 people [26]. According to the Rapid Damage and Needs Assessment (RDNA) conducted in the aftermath of the explosion by the World Bank Group, in cooperation with the European Union and the United Nations, to inform the Beirut Reform, Recovery and Reconstruction Framework (3RF), the overall physical damages caused by the explosion were estimated to be between USD 3.8 and 4.6 billion, with the housing sector being the most adversely affected (USD 1.9–2.3 billion) [26]. The RDNA also estimated the total losses in economic flows to be between USD 2.9 and 3.5 billion, with housing being again the most severely hit (USD 1–1.2 billion), followed by transport and port (USD 580–710 million) [26]. A two-week state of emergency was declared on 5 August, and national and international search and rescue teams, medical professionals, and disaster management experts responded to support the emergency efforts [27].

There are already several studies that use remote sensing to estimate damage associated with this event. Agapiou [28] uses both SAR and optical openly accessible European Space Agency (ESA) Sentinel satellite data in parallel by comparing change detection between images taken before and after the explosion. For the SAR data the magnitude of the backscattered signals of the two images are compared, as well as change of coherence by means of the formation of an interferogram. Principal Component Analysis (PCA) was used to analyse the Sentinel-2 optical images. Pilger et al. [29] investigated three independent methods (seismological analysis, acoustic yield relations, as well as InSAR satellite image analysis) for damage loss estimation and found that the yield range was consistent across the methods by one order of magnitude.

3.1.1. Remote sensing data

The pre-disaster and post-disaster optical images were collected at high resolution (nearly 50 cm pixel resolution) by the WorldView-2 satellite (obtained on 31 July 2020 prior to the disaster and 5 August 2020 after it) and Sentinel-2 provided images at 10 m resolution (obtained on 24 July 2020 before the disaster and 18 August 2020 after it), shown in Fig. 2.

The SAR images were collected at high resolution ("stripmap" mode, approximately 3 m by 3 m pixels) by TerraSAR-X (obtained on 9 June 2020 before the disaster and 7 August 2020 after the disaster). Medium-resolution images were obtained from ESA's Sentinel-1 satellites. Sentinel-1 image pairs obtained on 30 July 2020 and 5 August 2020 (before and after the explosion) were used.

3.1.2. In-situ data

One of several initiatives undertaken in response to the blast was a rapid visual assessment at the building level, undertaken by the Municipality of Beirut and the Governor of Beirut with support from UN-Habitat Lebanon. Covering all building types within a 2 km radius of the blast, the assessment aimed to gain an understanding of the extent of damage and especially structural impacts, identifying buildings or building elements at risk of collapse and in need of evacuation in the immediate term while also informing the formulation of early recovery measures. Data was gathered by field surveyors from several volunteering engineering consultancy firms starting immediately after the explosion until 11 September 2020. Based on the collected data, buildings were classified into different categories, depending on their damage level and implications/actions required (Fig. 3) [30].The database and Geographic Information System (GIS) put together for this initiative was used as in-situ validation data for input or validation (as appropriate) to compare with the remote sensing estimations.

A similar rapid building-level assessment was also conducted – with support from UN-Habitat Lebanon – by the Municipality of Bourj Hammoud, which is located in close proximity to Beirut Municipality and was also highly affected by the explosion [31].

3.2. Haiti earthquake, 2021

On the morning of 14 August 2021, a magnitude 7.2 earthquake occurred in Haiti. This differed significantly from the similarly powerful one that made Haiti headline news in 2010 (which affected the densely populated capital, Port au Prince, with an estimated death toll reaching 220,000). The epicentre for the recent event was in the more rural southwest region of the country. Although a few significant towns were affected, the total number of fatalities was around 2200. There was, however, widespread damage to buildings, and a response effort was needed over a large area. Results from early inspections allowed the buildings that were in a safe condition to return to normal usage with considerable benefit to the community.

Following the event, the Haitian authorities along with the United Nations and a number of international bodies, requested the activation of the International Charter on "Space and Major Disasters". Satellite data over the region was made available, free of charge, with the

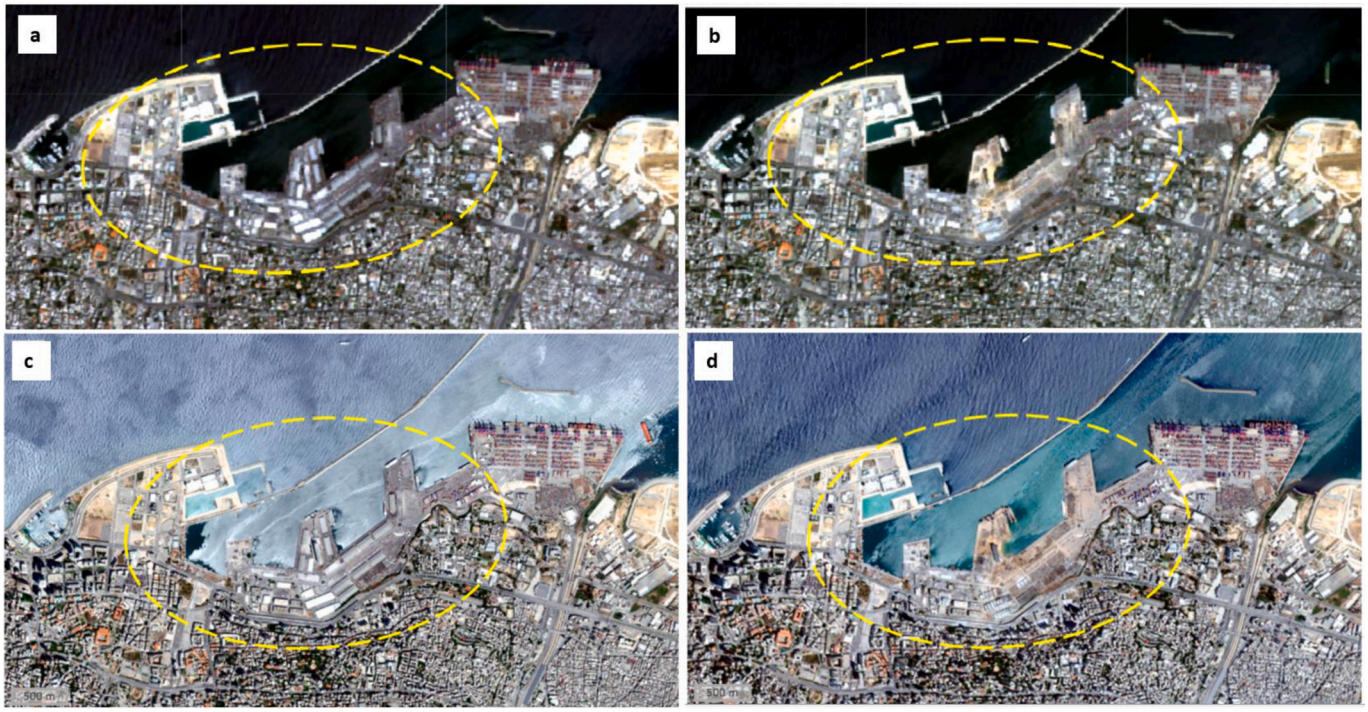

Fig. 2. Optical imagery from Sentinel-2 and WorldView-2 acquired over Beirut before and after the explosion. a) Sentinel-1 prior to the disaster (24 July 2020); b) Sentinel-1 after the disaster (18 August 2020); c) WorldView-2 prior to the disaster (31 July 2020); d) WorldView-2 after the disaster (5 August 2020).

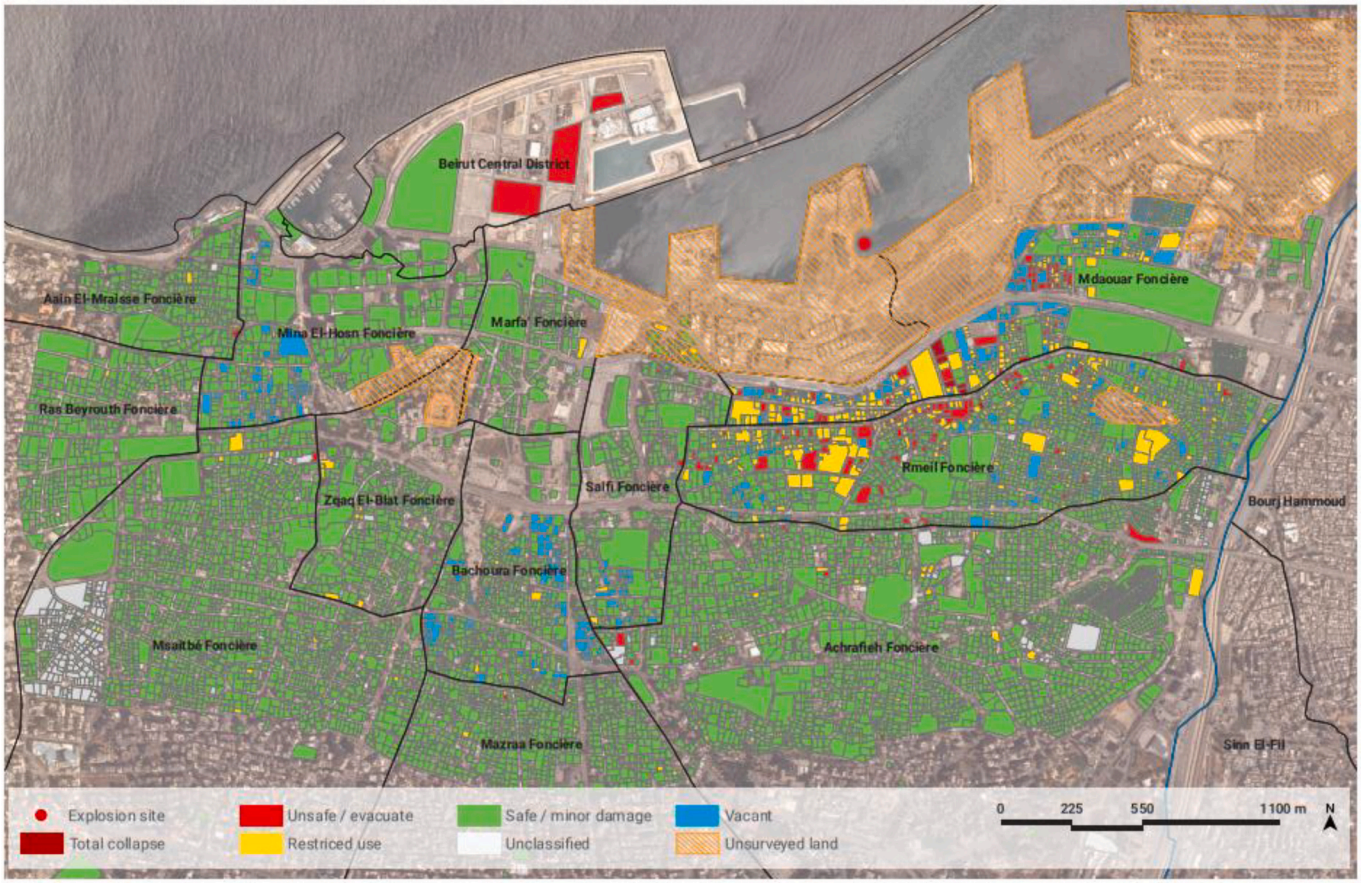

Fig. 3. Detailed map showing damage categories of assessed plots as at 5 September 2020, as shown in [30]. Reproduced with permission of UN-Habitat Lebanon.

international space community acquiring satellite imagery as quickly as possible. A number of different comparison maps were generated using very high resolution optical and radar sensors. Preliminary building damage assessment was conducted by the United Nations using very high-resolution satellite imageries pre and post event, through visual interpretation by the United Nations Satellite Centre (UNOSAT) and partners [32]. Other sources provided additional information related to the hazard, for example, ground motion maps ([33,34]. Whitworth et al., [35], as part of the Earthquake Engineering Field Investigation Team (EEFIT), carried out a hybrid mission for damage assessment and explored using satellite imagery as part of a process to undertake remote structural damage assessments in regions affected by the earthquake.

3.2.1. Remote sensing data

The pre-disaster and post-disaster optical images were collected at high resolution, and were provided by Maxar's Open Data Program. The resolution available was 55 cm or better in resolution, but was resampled to 55 cm resolution to obtain consistent resolutions. The images for Jérémie were collected on 3 January 2020 (before the earthquake) and 15 August 2020 (after the earthquake); for Les Cayes 27 February 2020 (before) and 15 August 2020 (after); for Miragoâne 14 August 2021 (after only); and for Port Salut 17 December 2020 (before only).

3.2.2. In-situ data

Damage and repair assessments were initiated by the Haitian Ministry of Public Works (MTPTC). These were conducted across the affected regions south of the island in partnership with the United Nations Office of Project Services (UNOPS), with Miyamoto International as technical partners. More than 550 Haitian engineers and social communicators were trained and sent to the affected areas to perform assessments, assessing 179,861 buildings over five months. This included diverse and challenging terrain, such as dense urban areas, steep mountains, and jungles. The database and GIS collection put together for this initiative was again used as in-situ validation data for input or validation (as appropriate) to compare with remote sensing estimations.

The Haiti analysis additionally makes use of building features collected during the structural damage assessment, to simulate a case where a region has surveyed and collected information about the region's building stock. Features that were used in this study were building area, number of floors, building age estimate, floor material and whether buildings were made with traditional building types/materials or with confined masonry.

3.3. Türkiye earthquake, 2023

South-east Türkiye and north-west Syria were struck by a magnitude 7.8 earthquake on the early hours of 6 February 2023. This was followed by a second earthquake of magnitude 7.5 just nine hours later [34]. Thousands of aftershocks continued to affect the area in the subsequent days and weeks. According to the World Bank, more than 41,000 fatalities, 108,000 injured people and 1.2 million people displaced were reported in the two weeks after the earthquake [36]. In the hours and days immediately after the earthquake, local first responders deployed (in the first hours), followed by international USAR teams (in the days after) that were deployed to assist local emergency services with the rescue efforts.

3.3.1. Remote sensing data

The pre-disaster and post-disaster optical images were collected over a number of different cities in Türkiye at high resolution (WorldView-2 satellite, at nearly 30 cm pixel resolution), and they were provided by Maxar's Open Data Program. The images collected for Adiyaman were from 18 January 2023 (before the earthquake) and 8 February 2023 (after the earthquake); for Antakya, 22 December 2022 (before) and 8 February 2023 (after); for Gaziantep, 2 January 2023 (before) and 8 February 2023 (after); and for Islahiye, 27 December 2022 (before) and 8 February 2023 (after).

3.3.2. In-situ data

A full collection of building damage classification data was not available for this study, and so a validation dataset was created using manual labelling of damaged buildings on high-resolution satellite data. As such, there were only two damage classifications: no visible damage and severe/total collapse.

4. Methods

In order to enhance the information provided to USAR operational organizations when responding to a disaster event, a data analysis architecture must fulfill a few crucial operating and usability conditions, specifically:

- It has to be very light in terms of power consumption, so to make the best use of the computational power available to a USAR organization in the field. In fact, it is expected that the architecture proposed in this manuscript could be employed by USAR units on ground, so to improve the timeliness of the response. In this case, it is worth remembering that USAR units on ground typically have little computational power (e.g., a laptop with no GPUs to preserve power consumption and work with limited data connectivity) and limited or scarce experience in data analysis;
- The architectures to be used by USAR units on ground must be easy to interpret and use. Indeed, it is important to point out that these units typically show limited extent of data analysis experience (the USAR units on ground typically consist of USAR technicians, construction experts and structural engineers, paramedics and medical experts, and some teams include volunteers from varied backgrounds; especially in early response situations, prior to the UN and international deployments, these functions may be local emergency services). Therefore, it is not reasonable to expect these units to be able to deal with the sophistication of the various steps for the set-up of complex systems for data analysis such as the deep learning-based architectures (e.g., hyperparameter tuning, design choices such as number of neurons, number of layers, activation functions, etc.). This aspect becomes even more dramatic when considering the stream of information that becomes available as time goes by, as sophisticated (with respect to the USAR units experience) architectures would require retraining and multiple runs to identify the most reliable model for the task.

With this in mind, it is possible to appreciate that the modern technologies used in academic literature for damage assessment are often not suitable to be applied in field operation scenarios. The most recent developments in data-driven damage assessment systems (in particular relying on remote sensing records – see for example [17]) typically rely on sophisticated architectures for data processing. These are typically based on identifying hidden patterns in the considered datasets by means of neural networks models. However, the reasons that such architectures are not typically suitable for field scenarios are the challenges posed by the characteristics of the data available for damage assessment, i.e., semantics unbalance (e.g., the distribution of damaged and undamaged assets is very often not uniform and indeed very skewed in one of the two directions), and limited records available for model training [17,32,34,35]. Indeed, the various satellite platforms and imaging techniques available capture different characteristics and properties of what is occurring on the ground. Combining these multiple modes would enable us to gain a more comprehensive understanding. However, information contained within these disparate sources is complex and non-linear, and the combination of multimodal datasets raises several challenges [2,13]. Approaching these technological challenges by means of systems based on convolutional neural networks [17]

have proven to achieve high accuracy in their predictions.

However, this typically comes at a structural and computational cost, i.e., the architectures are implemented on platforms that rely on GPU data processing, large scale data storage for model parameters update and high availability of power for computations [17]. These conditions are hence extremely hard to be matched in field operations for damage assessment (e.g., a USAR contingent on ground can take advantage of a laptop with no GPUs, 4 CPUs and less than 32GB RAM with extremely scarce access to power supplies), which strongly limits the applications of these modern technologies in operational scenarios. It is therefore important to consider these factors in the design of architectures that can be used to actually enhance the information processed by USAR organizations, so to achieve accurate and reliable information extraction whilst not hindering the actual use of these technologies in operational scenarios [32,34,35].

Graph representations [37] are a popular means of representing the network structure of connected data. The data are modelled as a set of nodes on a weighted graph, created by forming edges between nodes to represent the underlying structure and relationships between nodes (Fig. 4).

Methods selected for this context need to be flexible, fast and scalable in their deployment, as described in the operational requirements in Section 2. Graph-based data analysis methods can aid in such scenarios. Graphs, compared to other data learning methods, can be extended and updated as additional information becomes available, without needing to rebuild new models or data architectures. This additional information could include additional remote sensing sources, but also data collected in-situ or other contextual information provided as the response evolves. Such methods also allow for damage probability or uncertainty to be assigned to each node of the graph as part of the analysis – and these nodes and uncertainty estimations can also then be updated as additional input data becomes available. In this work, we look at the application of two methods that can create a framework to learn from sparse data sources and provide flexibility for input across the operational timeline. As more information becomes available, the output of the estimations can be improved. If this is combined with uncertainty estimation, the resulting system can be used by end-users as part of their decision-making at any given moment, and updated as more information is received from the ground. The flowchart in Fig. 5 shows on a time scale how the two steps of the proposed architecture are used and how the different components are used in the critical timeline of the post-event USAR operations.

In the immediate aftermath of a disaster, a method that works quickly with limited quantities of information will provide initial estimates when the entire ground context is unknown. With this, a quick means of providing a rough estimate of uncertainty may be useful if it can be provided just as quickly. Once this first rough estimate is calculated, a more refined estimate can be calculated as each dataset is analyzed, or the in-situ damage assessment information is delivered to the coordination point (provided that it too can be calculated quickly and repeatedly). This information is needed immediately - in the order of minutes rather than hours or days of computation. Furthermore, as more information is fed into the analysis, this analysis will need to be updated, also in the order of minutes, to provide the required information in a relevant and timely manner. In this paper we first look at belief propagation to provide a rough estimate, and then uncertainty-aware graph convolutional networks to provide an estimation of uncertainty.

4.1. Classic belief propagation framework

In the immediate aftermath of a disaster, there may be little to no ground information. Depending on the location and nature of the disaster, communication and transport infrastructure (key roads, bridges, tunnels, etc.) may be damaged or limited and there may be little information to go on prior to sending reconnaissance, USAR and emergency teams. If some remote sensing information was available, this could support authorities with the prioritization of resources and activities.

A simple, quick and robust means of quantifying the uncertainty would be to represent the data acquired on the region of interest in terms of a graph, and use this platform to extract information on the properties of the areas affected by the given disaster. Specifically, graphs are

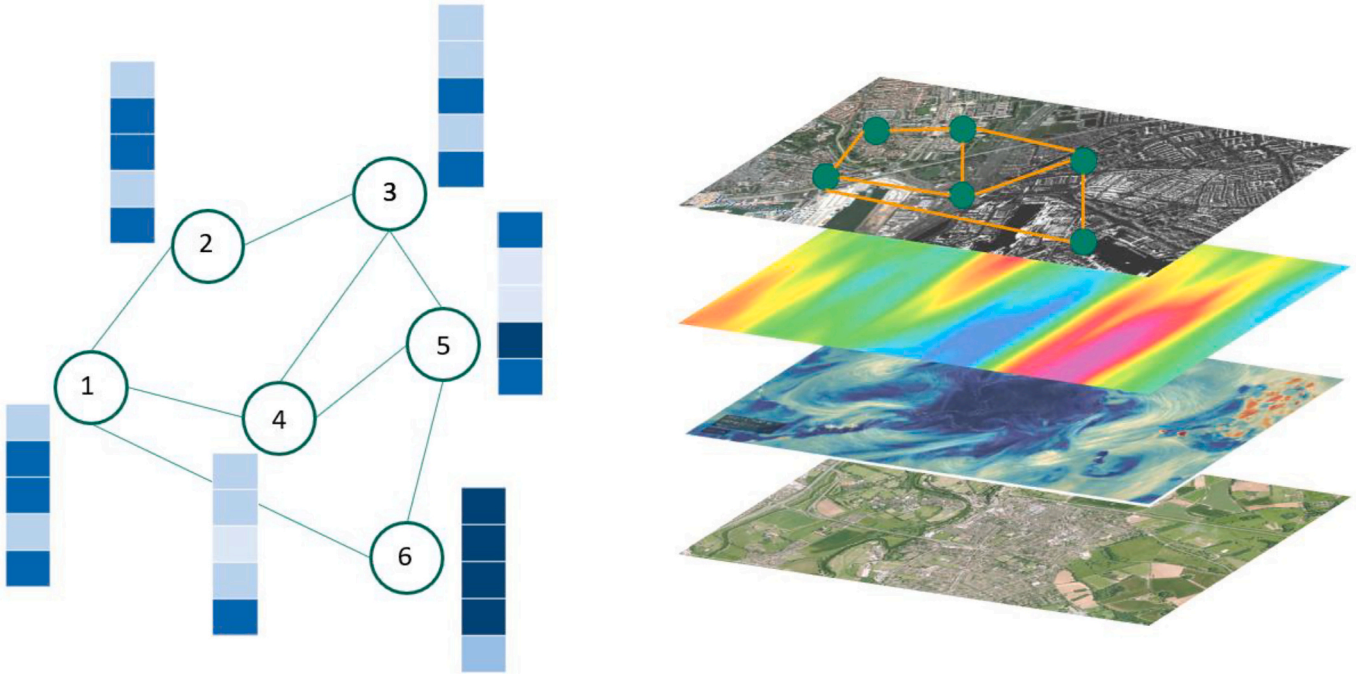

Fig. 4. Graphs consist of two main components: nodes and edges. Each node (circles in the left hand side diagram) identify one sample in the given dataset, and is characterized by a feature vector (represented by an array of blue shaded boxes in this figure). The weight associated with each edge (line connecting a pair of nodes in this diagram) models the similarity between the nodes under exam. In our case, nodes can be associated with each location for which pixel or building (see diagram on the right hand side). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

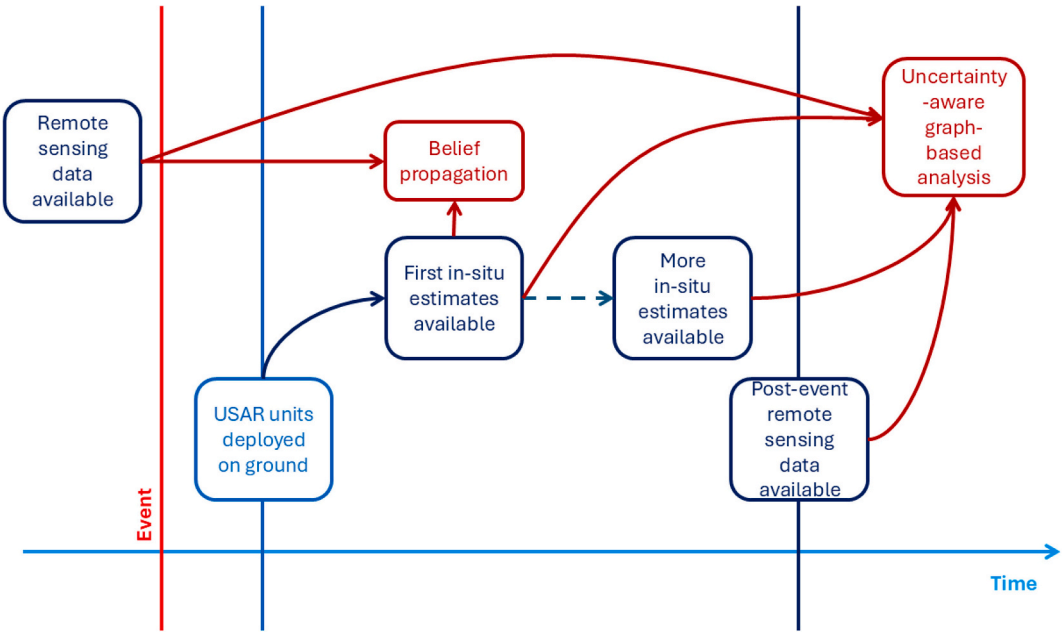

Fig. 5. Methodological framework detailing the main steps of the proposed architecture from across the timespan of an event and response. In order to provide timely initial response for USAR units on the ground, the proposed methodology consists of two steps. The first considers integrating pre-event datasets available with preliminary estimates from units on ground (belief propagation). Once more data is more available (either in-situ estimates of remote sensing), the uncertainty-aware graph-based analysis will be used to complement the preliminary estimates provided by the belief propagation algorithm and provide a complete overview of the characteristics of the impact of the disaster being considered.

mathematical structures that consist of two elements: nodes and edges. In data analysis, the nodes are used to represent the samples in a dataset. Hence, each node in the graph is characterized by the features that are associated with the corresponding sample. The edges are instead used to model the similarities between samples, especially when a weight is associated with each edge. The characteristics of the graph are summarized by an adjacency matrix. Specifically, for a graph consisting of N nodes, we can write a matrix $\boldsymbol{A} = \{A_{ij}\}_{(i,j)\in\{1,\dots N\}^2}$, where the element $A_{ij} \in [0,1]$ identifies the similarity between node i and node j. At this point, navigating nodes and edges translates to mainly performing matricial operations that leads to understanding the hidden patterns and complex interactions between samples.

When applied to post disaster scenarios, the nodes of a graph represent geographical locations in the region affected by the catastrophic event. In urban scenarios, each node can identify a specific structure, e.g., each building in the area under examination. Among the graph-based algorithms that have been proposed in technical literature to extract information from graph structures, belief propagation plays a key role. This algorithm is especially useful when little information about the considered samples and their semantics (e.g., classes, labels) is available. Belief propagation takes advantage of the ability of graphs to provide information on the structure of the data (by the network generated by the set of edges across nodes). By translating the information propagation in terms of message passing operations across the nodes, belief propagation realizes an approximate inference process. Specifically, belief propagation algorithm assumes to translate the probability that node i belongs to class k in the k-th element of a C-dimensional vector associated with node i, $\boldsymbol{b}_i$, where C is the number of classes or thematic clusters in the considered dataset. This vector is called belief of node i. The main idea of belief propagation is that the belief of each node would evolve depending on the messages that it will receive from its neighbors, i.e., the nodes that are connected by edges with non-zero weights (the elements of the matrix $\boldsymbol{A}$) to it. In this vision, a message $\boldsymbol{m}_{ij}$ encodes the belief that node j has about what class node i should be associated with. Iterating this process, it is possible to achieve exact marginal probability distributions for each node, so to enable the interpretation of their properties.

Belief propagation within the graph structure allows for inference of damage based on the similarities between nodes [38]. Beliefs are propagated along edges linking node pairs according to their similarity in (i) input data values or (ii) geographical proximity of corresponding areas. Belief propagation mechanisms have historically been used for image rendering and classification. Images are represented as large and highly connected graphs, with nodes encoding pixel values [39]. The approach of label uncertainty propagation was later developed into confidence-aware algorithms for belief propagation [40,41], in fields outside of image classification. Work by Eswaran et al., (2017) [42] shows a demonstrably scalable and efficient algorithm for propagating label uncertainty. Here we are proposing to adapt this form of method for damage assessment classification through belief propagation quantifying label uncertainty through Dirichlet Multinomial distributions.

In this application, belief propagation is used to make use of small amounts of information received on ground, together with what can be identified by the remote sensing data collected, and propagates the known information to areas where we have very little understanding. Specifically, let us consider to collect in a dataset $\boldsymbol{X} = \{X_{il}\}_{(i,j)\in\{1,\dots N\}\times\{1,\dots L\}}$ the L features that have been acquired for each of the N data samples. In this case one sample is an individual 10x10m grid square on the map of Beirut. The adjacency matrix $\boldsymbol{A}$ is constructed by computing the similarity between each pair of nodes according to their distances with respect to each of the L input features. Thus, the belief propagation algorithm provides information on the probability of each node (as previously mentioned, a node is associated with each sample, hence a small grid square of the city in our case) to have been affected by the disaster to a high or low degree, quantified by a high or low damage level, respectively.

In particular, if we consider that the set of nodes for which some initial information is available on the damage level is grouped in the $\boldsymbol{\Lambda}$ set of nodes, and that the (k,t)-th element of a matrix $\boldsymbol{\Theta}$ identifies the affinity between classes k and t (being $\boldsymbol{\Theta}$ a C × C matrix, according to the notation we introduced previously), the main steps of belief propagation can be summarized as follows:

1. The belief vectors for the nodes in $\mathbf{\Lambda}$ are initialized according to a Dirichlet distribution drawn over the features and the a priori information of node i to be associated with each class $\widetilde{\boldsymbol{b}}_i$, i.e., $\boldsymbol{b}_\lambda = Dir\left(\widetilde{\boldsymbol{b}}_\lambda; \widetilde{\boldsymbol{b}}_\lambda + \mathbf{1}\right) \propto \prod_{s=1}^{C} \widetilde{b}_s^{\widetilde{b}_{ls}-1}$, for $\lambda \in \mathbf{\Lambda}$. This choice is motivated to avoid the propagation of redundant information across the graph. Let us store this information in another set of vectors $\boldsymbol{e}$, i.e., $\boldsymbol{b}_i = \boldsymbol{e}_i$;
2. The messages between nodes are initialized to the similarity between nodes, i.e., a function of the distance between belief vectors;
3. The belief of each node in the dataset is updated as $\boldsymbol{b}_i \longleftarrow \boldsymbol{e}_i + \sum_{j\in\Gamma(i)} \boldsymbol{m}_{ji}$, being $\Gamma(i)$ the neighbourhood of node i, and $\boldsymbol{m}_{ji} \longleftarrow \mathbf{\Theta}\left(\boldsymbol{e}_j + \sum_{z\in\Gamma(j)\setminus i} \boldsymbol{m}_{zj}\right)$;
4. Iterate point 3) until all the nodes have been assigned a belief vector.

At the end of this process, by applying a probability threshold to the belief vectors associated with nodes corresponding to buildings, it is possible to derive an initial estimate of the level of damage of each building in the considered region, hence obtaining a fast and consistent assessment of the level of damage each building has suffered. Thus, this method is a means of addressing efficient solutions given very limited data.

4.2. Uncertainty-aware graph-based framework

To quantify the uncertainty in our predictions (i.e., how sure are we that the damage assessment estimates are correct) we can assess how well different measures of model uncertainty perform with respect to detecting points that are out of distribution from distributions observed during initial training. Here we do this by adopting graph-based neural network architectures that are adapted to provide subjective opinions [43] through the use of Dirichlet distribution parameterizations [44,45]. Subjective logic [46] can be deployed to measure quantities of belief based on different criteria:

1. Uncertainty due to lack of evidence, or 'vacuity', with a higher value of vacuity suggesting a lack of supporting evidence for a prediction; and
2. Uncertainty due to the presence of conflicting information, or 'dissonance'.

The uncertainty-aware approach [43] (an adaptation of the graph convolutional neural network architecture [45] is a method for obtaining classification predictions with a quantification of confidence. By adapting the neural network's output activation function and selecting an appropriate loss function during training, the model can be trained to provide the concentration parameters which characterize a Dirichlet probability distribution [41]. The Dirichlet distribution can be considered a second-order probability distribution in the sense that it is a distribution of probabilities over multinomial probability distributions. The concentration parameters of the Dirichlet distribution, which the model is trained to provide as outputs, can be mapped to measures of uncertainty.

The implementation of the uncertainty-aware graph-based neural network in this study uses graph nodes to represent individual buildings and performs a semi-supervised classification task. The classes represent building damage categorisations; semi-supervised in this context means that a subset of building damage labels are provided during training and the model is asked to classify the building damage of the remaining buildings. This represents the scenario that would exist in the immediate aftermath of a disaster when some buildings have been assessed and we want to consider whether we can make inferences about the damage extent of other buildings using this information. In order to make these inferences, the model is able to make use of the existing damage labels as well as high-resolution pre-event and post-event optical imagery. Where available, this can be expanded to include information about the building, including information about its construction, height and age. The structure of the graph is designed to reflect similarities between buildings and so edges between buildings are computed so as to connect buildings to one another based on their proximity in this feature space.

Specifically, we use dataset $\boldsymbol{X} = \{X_{il}\}_{(i,j)\in\{1,\dots N\}\times\{1,\dots L\}}$ to represent the L features that have been acquired for each of the N buildings represented by graph nodes. The uncertainty aware approach trains the neural network so as to fit a function $\boldsymbol{\alpha} = f_\theta(\boldsymbol{X})$, where $\boldsymbol{\alpha} = \{\alpha_{il}\}_{(i,j)\in\{1,\dots N\}\times\{1,\dots L\}}$ and θ represents the parameters of the neural network. Each row in the matrix $\boldsymbol{\alpha}$ represents the concentration parameters of a Dirichlet distribution. Ordinarily in a classification task the softmax activation function is used on the output layer of the neural network so as to obtain outputs which are non-negative and sum to 1 such that they can represent probabilities. To obtain outputs which can represent concentration parameters a different output activation function must be applied, in this case $\alpha_{il} = 1 + max(0, z_{il})$, where α_{il} and z_{il} represent the output and input to the layer respectively. This activation function ensures that $\alpha_{il} \geq 1, \forall i, l$. To relate $\boldsymbol{\alpha}$ to predicted classes and confidence levels we must first introduce y_l to represent the probability of a given sample belonging to the l^{th} class and therefore $\boldsymbol{y} \in \mathbb{R}^L$ as the vector obtained by stacking the predicted probability into a vector. The Dirichlet distribution is a probability distribution over all the possible $\boldsymbol{y}$ vectors. If the distribution is sharp, which means the probability density is concentrated within a small region, then we are confident about what $\boldsymbol{y}$ is. This is different to being confident about which class the sample belongs to, as the $\boldsymbol{y}$ vector which we are confident in might represent each class being equally probable. Alternatively, if the distribution is flat, which means the probability density is spread, then we are not confident about the $\boldsymbol{y}$ vector. The sharp distribution in the equally probable region of the possible $\boldsymbol{y}$ represents the case which would give high values of the quantity dissonance which we have described, whereas the flat distribution over $\boldsymbol{y}$ represents the case when the quantity vacuity would be high.

The separation of uncertainty into vacuity and dissonance allows the model to express why it is uncertain as well as when. A lack of confidence due to an absence of evidence, indicated by high vacuity, is likely to require a different interpretation than a lack of confidence due to observations of conflicting evidence, indicated by high dissonance. This distinction, which can be made due to the subjective logic model, is particularly well-suited to applications such as post-disaster damage assessment which requires a higher level of interpretability.

We obtain the measure of confidence provided by vacuity for the i^{th} sample according to the equation $\text{vacuity}_i = \frac{L}{\sum_{i=1}^{L} \alpha_{il}}$. The dissonance for the i^{th} sample is also a function of $\boldsymbol{\alpha}$ and we refer the reader to [43,41] for specific details.

In order to ensure that during training the model converges to a set of parameters θ which allow the neural network to predict $\boldsymbol{\alpha}$ which represents meaningful Dirichlet distributions for each node, the loss function which is minimized during training has been designed specifically for the uncertainty-aware approach [41]. It makes use of a component which uses a pre-trained teacher network as well as a Graph-based Kernel Dirichlet distribution Estimation (GKDE) prior.

5. Results

5.1. Use of belief propagation methods simulated over a response period

The belief propagation method was deployed on the Beirut dataset. To simulate the progression of the model's ability to make classifications during disaster response, a simulation scenario was set up. Starting with no information on the ground (i.e., just satellite information), we set up several scenarios in time where the algorithm was given a number of ground truth classifications as training prior to estimating the classification of the remaining buildings, to simulate on-the-ground

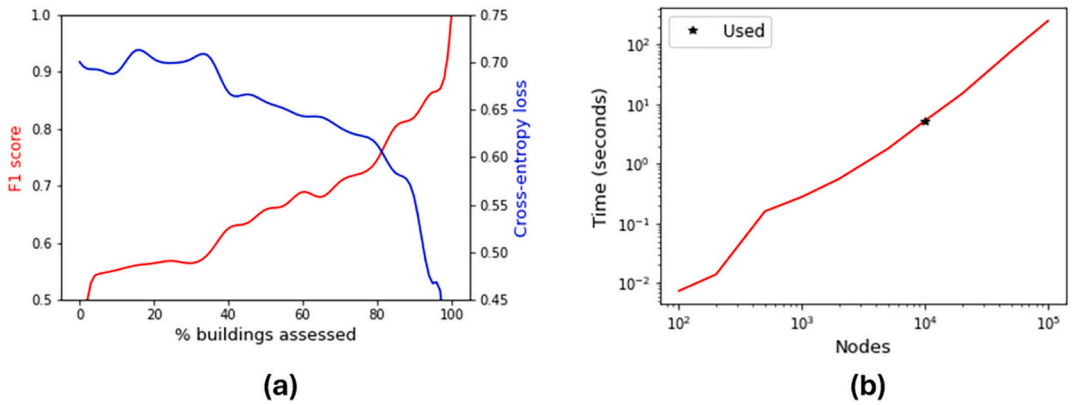

Fig. 6. a) Performance with increasing ground truth data quantified by the percentage of buildings assessed; b) Computation time with node increase. Evaluation performed with 6 edges per node on a standard laptop with 4 CPUs and 13GB RAM. If a node per pixel becomes too expensive, nodes can be sampled or assigned to larger patches. In this study, pixels were sampled when exceeding 10,000 nodes.

information returning from USAR teams and engineers. Fig. 6 summarizes performance evolution with increasing ground truth. The F1 score is used to evaluate the model accuracy, computing how many times a model makes a correct prediction across the entire dataset. It quantifies the ratio between correct predictions and misclassification.

The benefit of this method is that an estimate is immediately available, is updated as more information becomes available and an estimation of confidence is provided alongside damage classification estimates. This ability to update and produce an output for the entire area of interest is an improvement over state-of-the art estimations which are a snapshot in time of remotely sensed data, or final ground surveys. Rapid computation (Fig. 6b) and maximized data exploitation allow for frequently updated information (in any form) to increase performance as demonstrated.

However, In terms of classification reliability, Fig. 6a shows classification F1 score remains below 0.6 until 35 % of buildings are assessed and does not rise above 0.8 until 80 % ground-truthing. This unreliability could waste precious time for disaster response teams if allocating resources to a false damage classification and will decrease trust in the model.

To understand the practical implications of increasing accuracy with ground truth data availability, Fig. 7 shows the output predictions of models trained with 5 %, 25 % and 50 % of buildings assessed. With only 5 % of building assessment labels available, most damage label predictions remain too uncertain to be reliable estimates. However, as assessment coverage, and hence training label data availability, increases, a significant part of the city which is as yet unassessed is assigned an estimated damage label with a certain degree of confidence. Fig. 6 shows how the prediction of damaged and undamaged classes improves with increasing damage assessment coverage. The spatial distribution of the predictions is reported in Fig. 8.

It is worth putting these results in the context of using the proposed architecture in USAR field operations. In particular, we tested the proposed method and a supervised deep learning scheme proposed in technical literature [17] when employed over a computational platform typically in use by USAR units on ground, i.e., a laptop with 4 CPUs and 13GB RAM. Running multiple experiments (100 runs for three different setups of training data available, i.e., 5 %, 25 % and 50 %) and focusing our attention on the accuracy of the outcomes for identification of damaged and undamaged buildings, as well as estimating the required computational complexity of the two systems (quantified in terms of execution time), it is possible to obtain a thorough overview of the actual feasibility of the methods for use in operational USAR scenarios. These results are reported in Table 1. It is possible to appreciate that the most sophisticated scheme based on deep learning technology is indeed able to achieve higher accuracy. However, this comes at a computational cost that makes this type of approach unfeasible for USAR operational scenarios, where timely response (in order of minutes) are essential.

5.2. *Uncertainty-aware graph-based framework*

In the application context for this method, the nodes in the graph represent individual buildings. The node features used as model input are computed using statistical properties of the interferogram, the optical before and the optical after data within each building's footprint. In this case, the statistical summary included the minimum, maximum, mean, variance, skewness and kurtosis. An additional node feature was added which described the number of pixels within the building's footprint which was used as a proxy for building size).

5.2.1. *Beirut*

The most severe damage occurred at the site of the blast (marked with an "X" in Fig. 9), predominantly impacting the port and regions closest. Medium level classification (buildings marked in yellow for "restricted use") were mainly in the concentric ring out from the blast and red classified buildings, with the majority of buildings marked as green for "safe/minor damage". Fig. 9 shows that the adapted GCN method misclassifies a number of predictions around the blast site and to the east. However, these misclassifications are also flagged by the dissonance metric as being regions where the algorithm is highly certain about its predictions.

"Entropy" was also computed for each scenario for comparison. Entropy is a classic measure of uncertainty within the information engineering community. The issue with the entropy measure here is that it is an aggregate measure and only tells us that the data varies a lot from one point to another. The subjective logic that we deploy here (as vacuity and dissonance) aims to decrypt this entropy metric and get more details around the uncertainty that can be quantified, i.e. provide two measures that are more interpretable.

5.2.2. *Haiti*

Several regions in the impacted area of Haiti were analyzed using the adapted graph convolutional network with uncertainty framework. Fig. 10 displays the prediction output of the proposed modified approach in comparison with a more traditional GCN (as proposed by [45]). Uncertainty measures were computed using the "vacuity" and "dissonance" measures previously defined, and example outputs over the regions of Les Cayes and Jérémie are shown in Fig. 11.

5.2.3. *Türkiye*

Similar to the previous section, uncertainty measures were computed

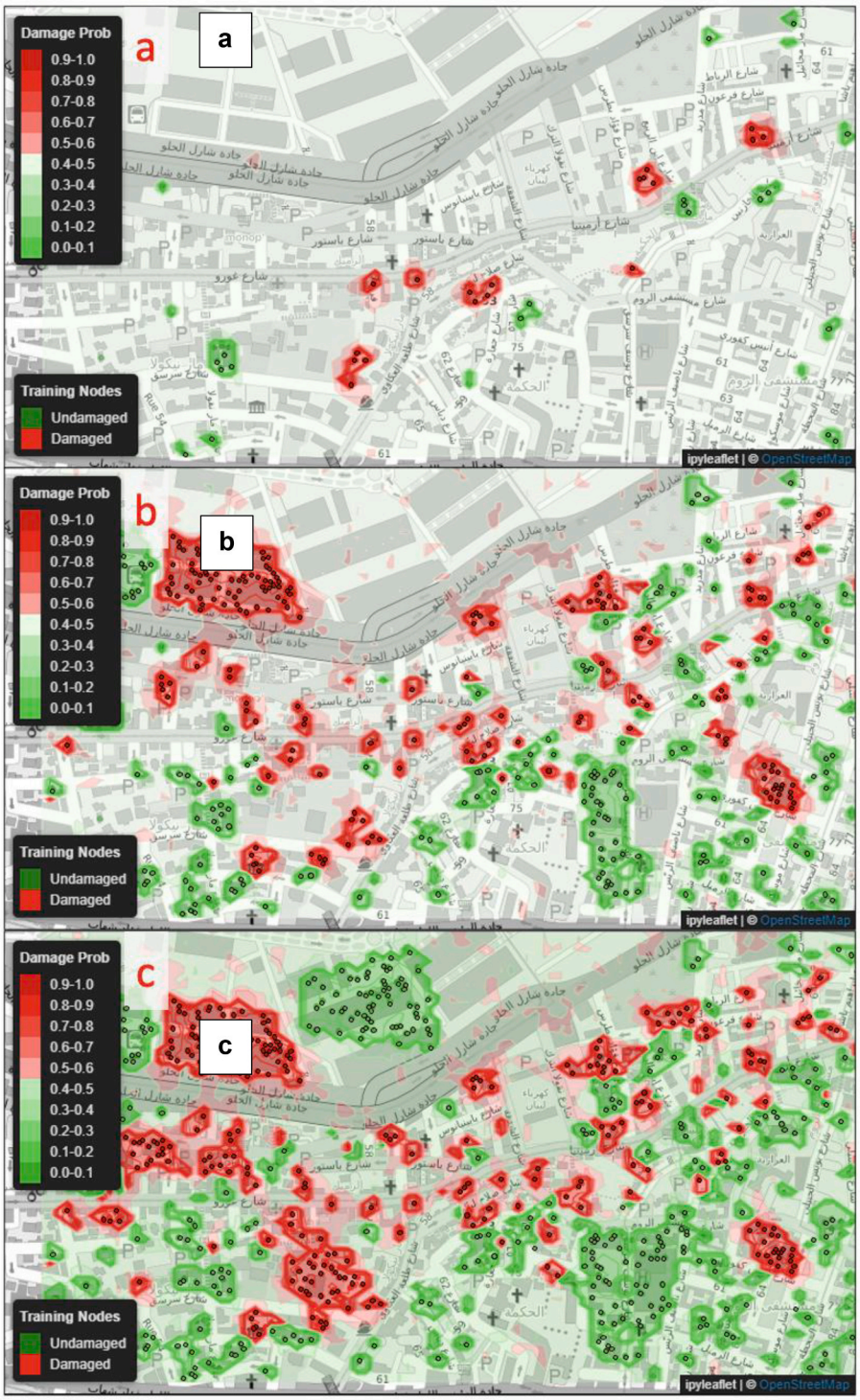

Fig. 7. Damage assessment prediction maps generated after a) 5 %, b) 25 %, and c) 50 % of buildings are assessed. The contours quantify uncertainty via a probability of damage. A 0.5 probability of damage indicates ignorance. The training nodes used are represented by circle markers.

using the "vacuity" and "dissonance" measures. Example outputs over the region are shown in Fig. 12.

6. Discussion

In this operational Search and Rescue scenario, there are several major challenges in trying to leverage remote sensing data. Many traditional data learning methods for classification and prediction are ruled out for practical implementation due to the very limited training data. With the limited data that is available, there may be further challenges, such as the need to investigate unbalanced datasets arising due to label noise, inaccurate labelling and missing information. In this work we investigate three real world case studies, each with their own challenges and specific dataset qualities. These different properties affect both the ability of data learning methods to make predictions, as well as the output of the uncertainty metrics that we have employed.

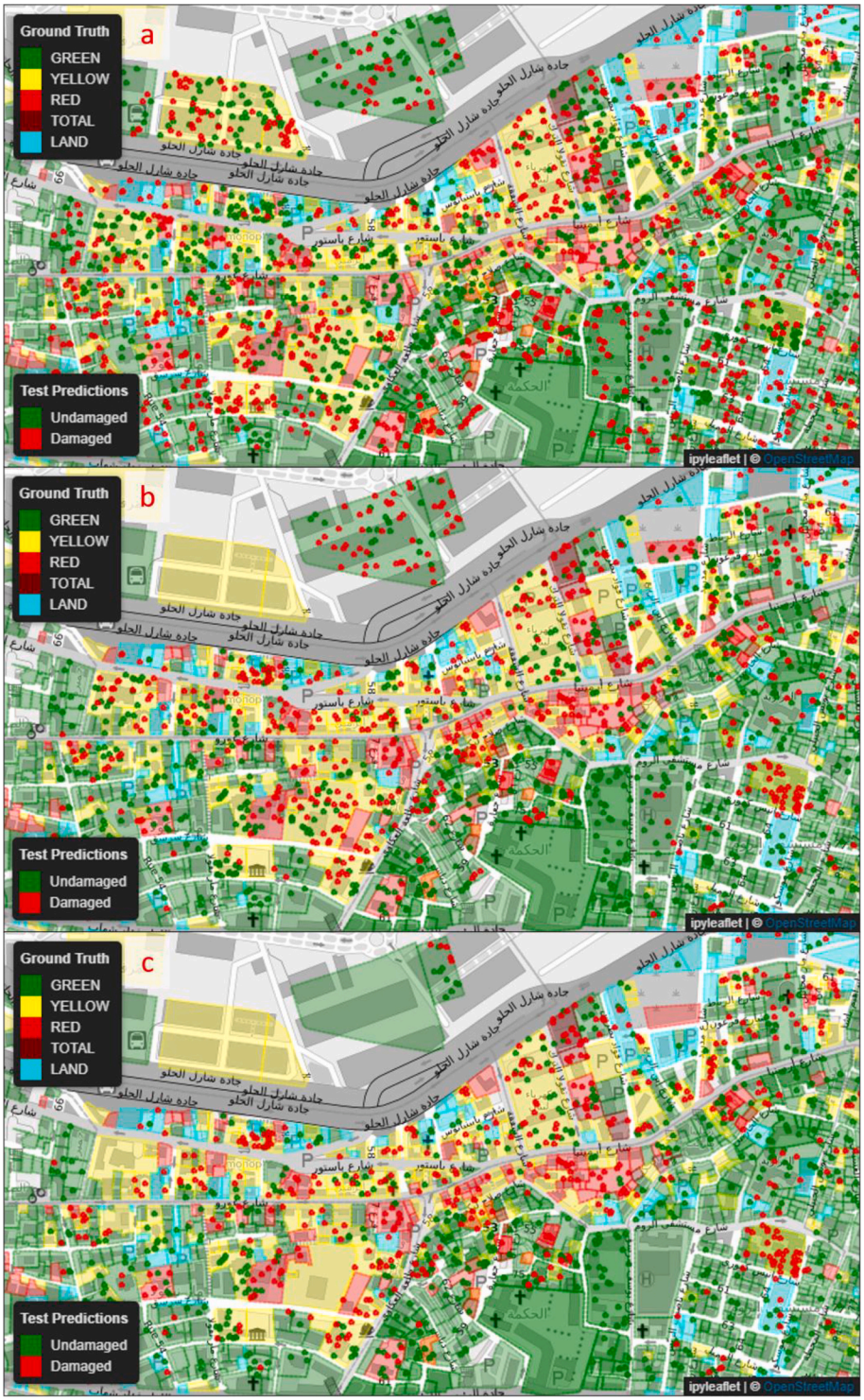

Fig. 8. Ground truth comparison with test node predictions as circle markers for a) 5 %, b) 25 %, and c) 50 % of buildings assessed.

6.1. Adapted GCN predictions

Methods based on adapting more classic belief propagation and GCN architectures were implemented in order to predict the damage classification, with varying degrees of success depending on the parameters of the specific case study. The developed method provides a means of efficient analysis with reasonable equipment for data processing, but more specifically it provides a means from which uncertainty can be conveyed to those making decisions based upon the damage estimation. In future, different methods for prediction could be employed, but understanding how much an operator can rely on the estimation is crucial in making decisions on the ground, for example, in deciding where to focus the efforts of specific reconnaissance or sectorization for sending resource. However, there are points to note specific to the adapted GCN method deployed across the three disaster scenarios.

For the Beirut case, we can to appreciate how the strong unbalance between the strongly damaged, mildly damaged and undamaged buildings significantly affects the classification prediction performance. However, it is also true that the proposed uncertainty metric shows higher values in correspondence with the pixels that have been

Table 1
Performance comparison of the proposed belief propagation scheme with state of the art algorithm proposed in Adriano et al. [17] under diverse conditions for training. Accuracy and execution time are reported when running the algorithms on a laptop with 4 CPUs and 13GB RAM, i.e. a typical platform for USAR units on the ground.

Training data	Accuracy [%]		Execution time [min]	
	Belief propagation	Supervised DL	Belief propagation	Supervised DL
5 %	65 ± 3	68 ± 2	20	112
25 %	67 ± 3	70 ± 3	23	165
50 %	72 ± 3	75 ± 2.6	24	186

incorrectly labelled by the graph convolutional network. Therefore, the proposed algorithm successfully achieved the goal of providing a reliable quantification of the confidence we can put on the estimated damage outputs.

With the analysis of the Haiti dataset, additional information regarding the building stock was available (albeit retrospective to the event. Including even a small, coarse amount of this contextual information would improve the damage predictions, highlighting the value and opportunity of collecting building information in advance of events. This could either be done manually through structural surveys of building stock, or through using methods that leverage remote sensing information to classify building by type, such as Geiß et al. [16] who then use this information to estimate vulnerability of buildings to hazards.

A very limited dataset was used for the damage analysis of the Türkiye earthquake. The validation dataset was created using manual labelling of satellite data, and not using a complete dataset of buildings surveyed by structural engineers on ground as was the case in the other sites. This limited the states of classification available meant that there were only two binary classes over a limited number of buildings to test for validation.

6.2. Uncertainty quantification

In order to properly quantify the ability of the proposed approach to provide useful information for the operational use of the aforementioned methods, we investigated the quality of the damage predictions and their associated levels of uncertainty. Specifically, we analyse the receiver operating characteristic (ROC) curves to show whether the uncertainty measures can be used to indicate confidence in the model's predictions. For each of the test sites considered in this paper, the ROC curves are derived by plotting the value of the true positive rate (or specificity, hence summarizing the accuracy of the classifier to identify the correct class of damage for the samples under exam) as a function of a threshold based on the false positive rate (i.e., the probability of false alarm). The area under the curve (AUC) provides the degree or measure of separability, i.e., it quantifies how much the model is capable of distinguishing between classes. Specifically, as much as the AUC value increases (it is limited between 0 and 1), the confidence we can put on the classification results increases too. Fig. 13 displays these curves for each test site.

We can appreciate how the measures of uncertainty that were proposed in Section 4.2 can help in digging deeper into the characteristics of the damage predictions made for each area under examination. In fact, the AUC we achieve by using vacuity and/or dissonance to run the prediction model is typically greater than (or at worst comparable to) the results that would be achieved by means of state-of-the-art techniques modelling uncertainty in the classification for damage assessment in all the considered test sites. Taking a closer look at the characteristics of these curves, we can further assess how the proposed approach can help in providing more details on areas under examination. To this aim, it is appropriate to consider the implicit meaning of the vacuity and dissonance metrics in terms of the type of uncertainty they can grasp.

Specifically, vacuity is a proxy for *epistemic* uncertainty, i.e., uncertainty caused by model parameters mismatch and non-ideal setup. As such, when the higher value of AUC is achieved when considering vacuity as uncertainty metric, we can assess that the main reason for the uncertainty in the model prediction can be caused by insufficient information or knowledge that the annotation process can use for generating the samples used for training. The extreme case of this uncertain condition would result in a uniform distribution of the likelihoods of a given sample to be associated with the classes under exam. In this case, to increase confidence in the predictions, using more sophisticated models (eventually based on nonlinear segmentation of the data space) or training samples with a finer granularity in terms of the classes to use

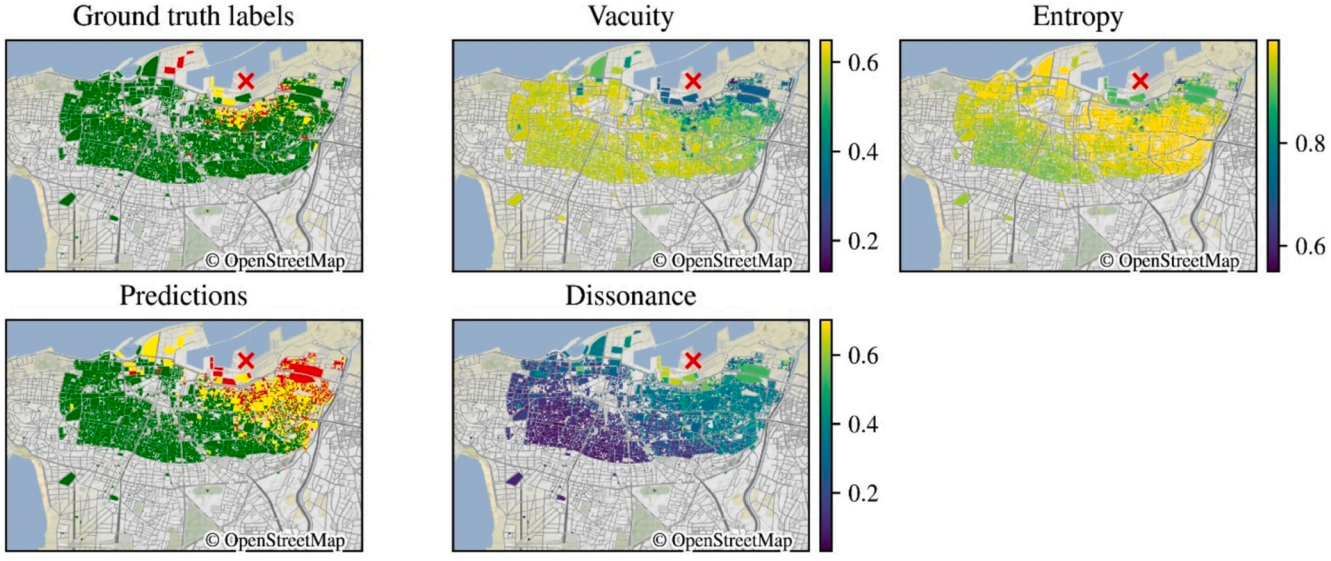

Fig. 9. Results of the adapted graph convolutional network (GCN) method and associated uncertainty quantifications for the Beirut data. Ground Truth displays the assessments carried out on the ground, with green for "safe/minor damage", yellow for "restricted use" and red for "unsafe/evacuated". Predictions are the estimations of these classes by the graph convolutional network using remote sensing data. "X" marks the blast location. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

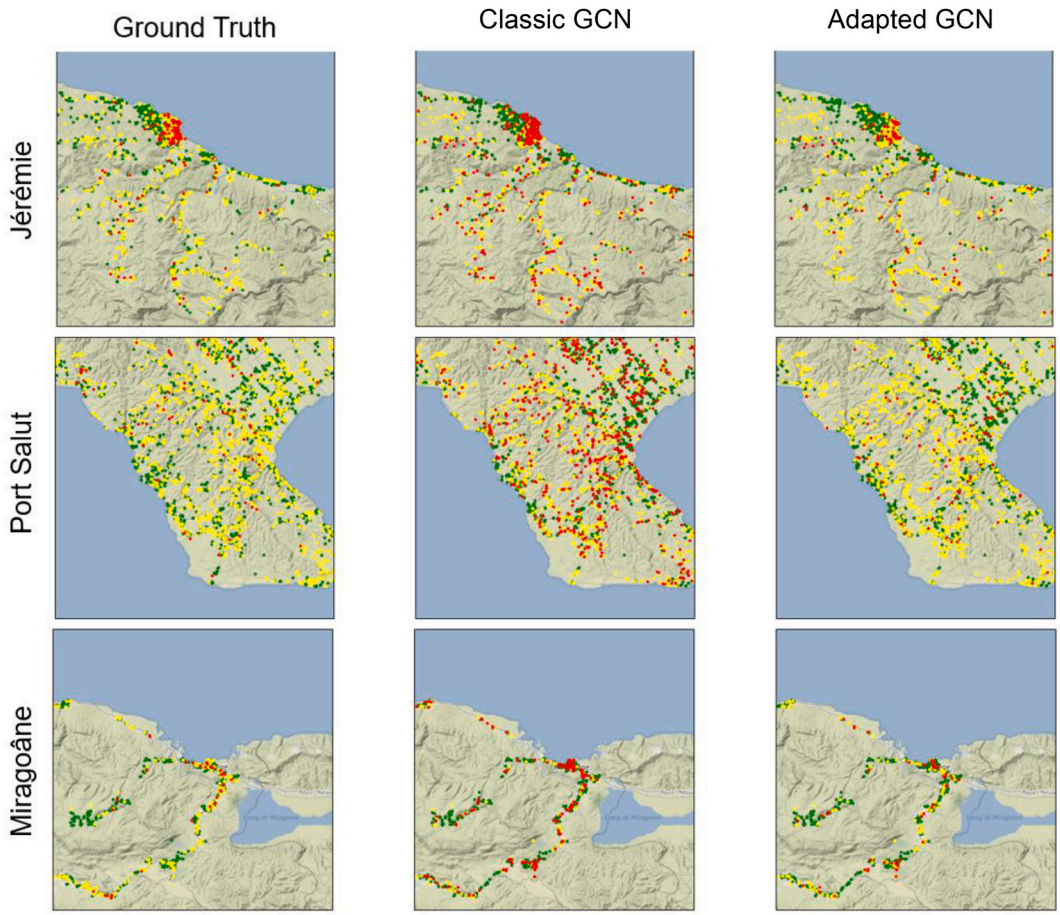

Fig. 10. Results of the adapted graph convolutional network (GCN) method compared with a classic GCN method. Ground Truth displays the assessments carried out on the ground, with green for "safe/minor damage", yellow for "restricted use" and red for "unsafe/evacuated". Predictions are the estimations of these classes by the graph convolutional network using remote sensing data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(e.g., increasing the number of classes for damage assessment) would be recommended.

Analogously, dissonance is a proxy for *aleatoric* uncertainty. In other terms, it measures the uncertainty that can derive from the variability in the data, leading to inconclusive decisions of the prediction model. Thus, when the higher value of AUC is achieved when considering dissonance as uncertainty metric, we can expect that conflicting evidences might arise in the datasets, leading to confused annotations. Pushing aleatoric uncertainty to the extreme would result in conflicting predictions, e.g., predictions assigning equal likelihood for classification to each of the considered classes. In this case, a deeper investigation of the quality of the data (eventually by means of adaptive feature selection [47] might result effective to increase the separability of the predictions, hence the confidence one can put in the analysis.

With this in mind, it is possible to fully appreciate the added value of the approach proposed in this work for uncertainty estimation of the damage assessment predictions in operational scenarios. Specifically, following the line of thought that we proposed above, we can notice that the state-of-the-art uncertainty quantification cannot provide the same level of details on the causes of uncertainty as the proposed method. In fact, entropy is a proxy of the sum of aleatoric and epistemic uncertainty [41]. As such, it cannot help in distinguishing what the real reason for uncertainty would be. On the other hand, it could be used to identify cases (such as the one in Fig. 11d) where to achieve confidence in prediction a deeper reconsideration of the data as well as of the annotations used for training should be put in place.

It is worth noting that this ability to provide a higher degree of granularity in identifying the causes of uncertainty results in tangible effects for the operational use of this method for damage assessment. If a higher value of AUC is obtained by considering vacuity, the end users could consider refining the definition of the classes for the analysis in specific areas, hence reconsidering the initial annotations to refine the analysis. On the other hand, when a higher value of AUC is achieved by considering dissonance, the end users could take into account the deployment of resources (e.g., drones) and personnel to collect additional information on the structures that have been hit by disasters. This would result in a substantial impact on the deployment of the operations for search and rescue in areas affected by disasters, as more precise decisions can be taken in these extremely stressful situations.

6.3. Performance comparison

Finally, analogously to the case of belief propagation, we compared the results we obtained when running the proposed algorithm on a platform typically in use of USAR units in the field with the outputs of a state-of-the-art system based on Bayesian neural network [48]. This performance comparison is reported in Table 2. It is possible to appreciate that the accuracy of the two systems do not significantly differ when used on the same platform available to USAR units. On the other hand, the execution time of the proposed approach is orders of magnitude less than the one required by the architecture in [48], based on a more sophisticated strategy of Bayesian deep learning. This result further emphasizes how the constrained computational power available to USAR units plays a crucial role in the actual working conditions of the

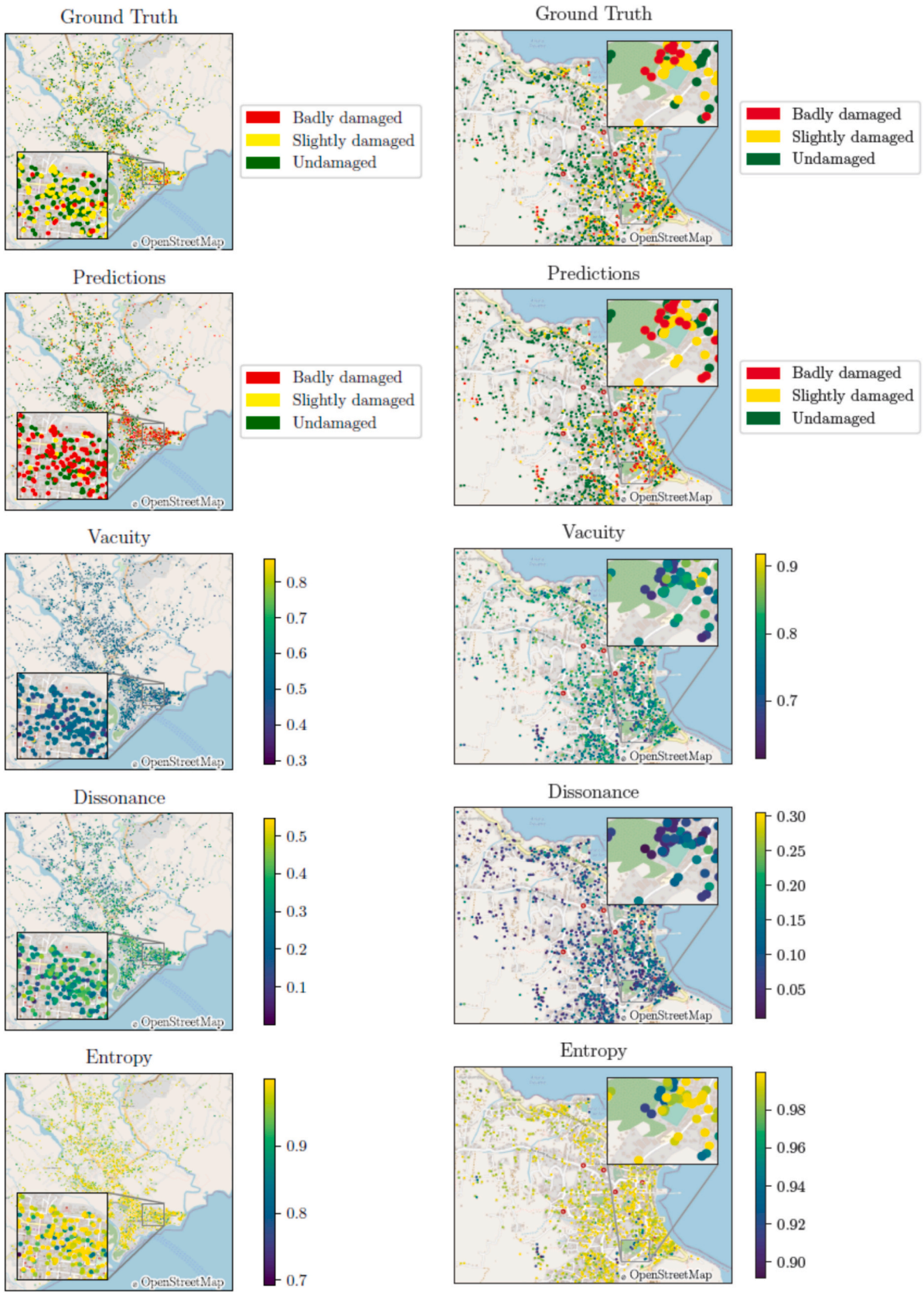

Fig. 11. Examples in closer view for two sites in Haiti, Les Cayes (left) and Jérémie (right). For each of these sites, the ground truth classification data is presented against the adapted GCN predictions, the vacuity and dissonance measures proposed in this paper, and calculated entropy values.

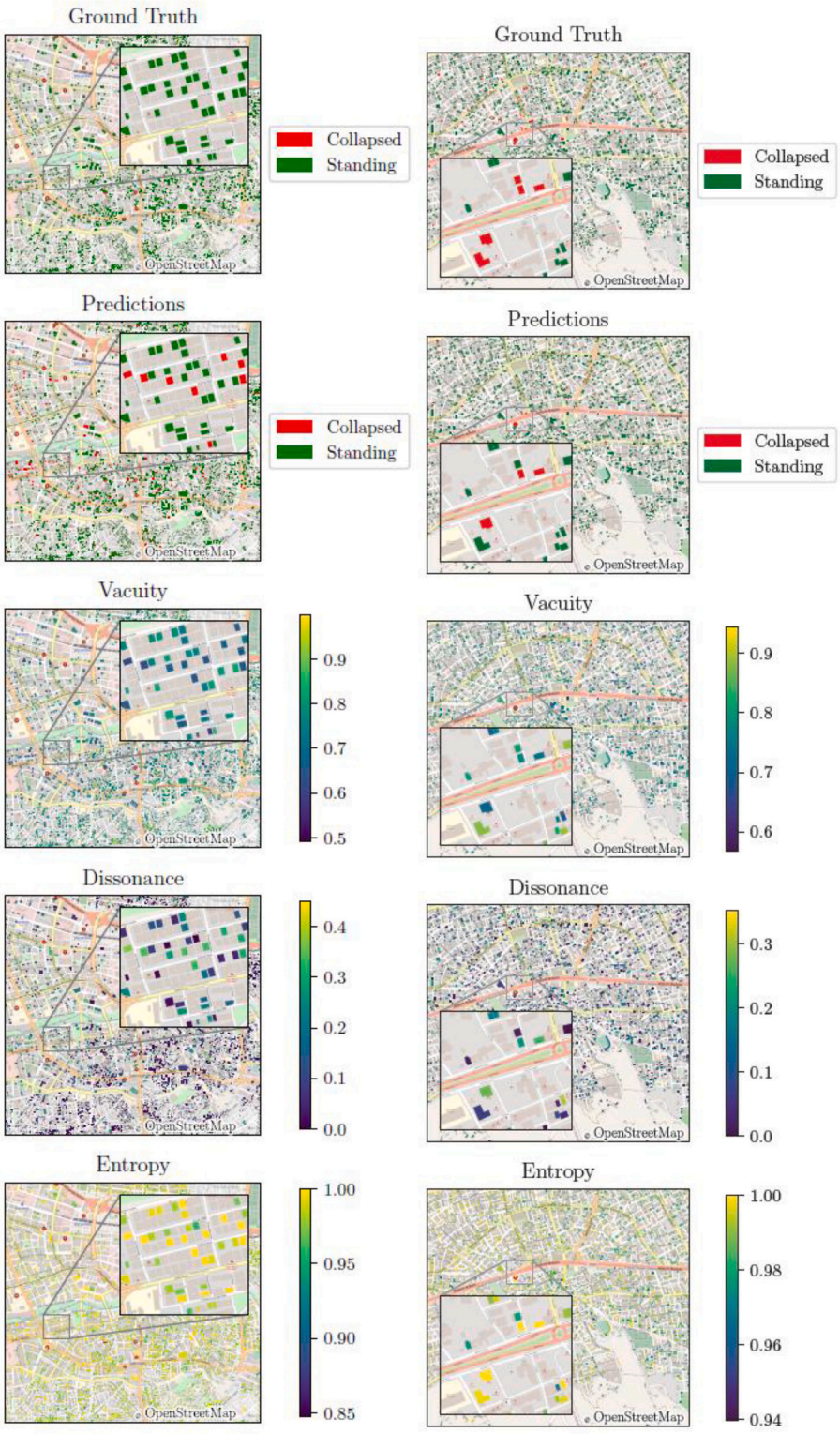

Fig. 12. Examples in closer view for two sites in Türkiye, Gaziantep (left) and Adiyaman (right). For each of these sites, the ground truth classification data is presented against the adapted GCN predictions, the vacuity and dissonance measures proposed in this paper, and calculated entropy values.

data analysis architectures to be used to support the USAR organizations. This comment is consistent with the results commented in the previous sections of this paper, highlighting the importance of ensuring good quality characterization of the damaged areas, together with timely responses and light data analysis architectures (i.e., reducing power consumption and able to work on platforms with limited computational capacity).

7. Conclusions

This work addresses the use of remote sensing for the specific operational context related to emergency USAR work following a disaster. To this objective, this work investigated two methodologies with two distinct goals. The initial approach involved a belief propagation technique, delivering prompt and robust outcomes based on limited initial

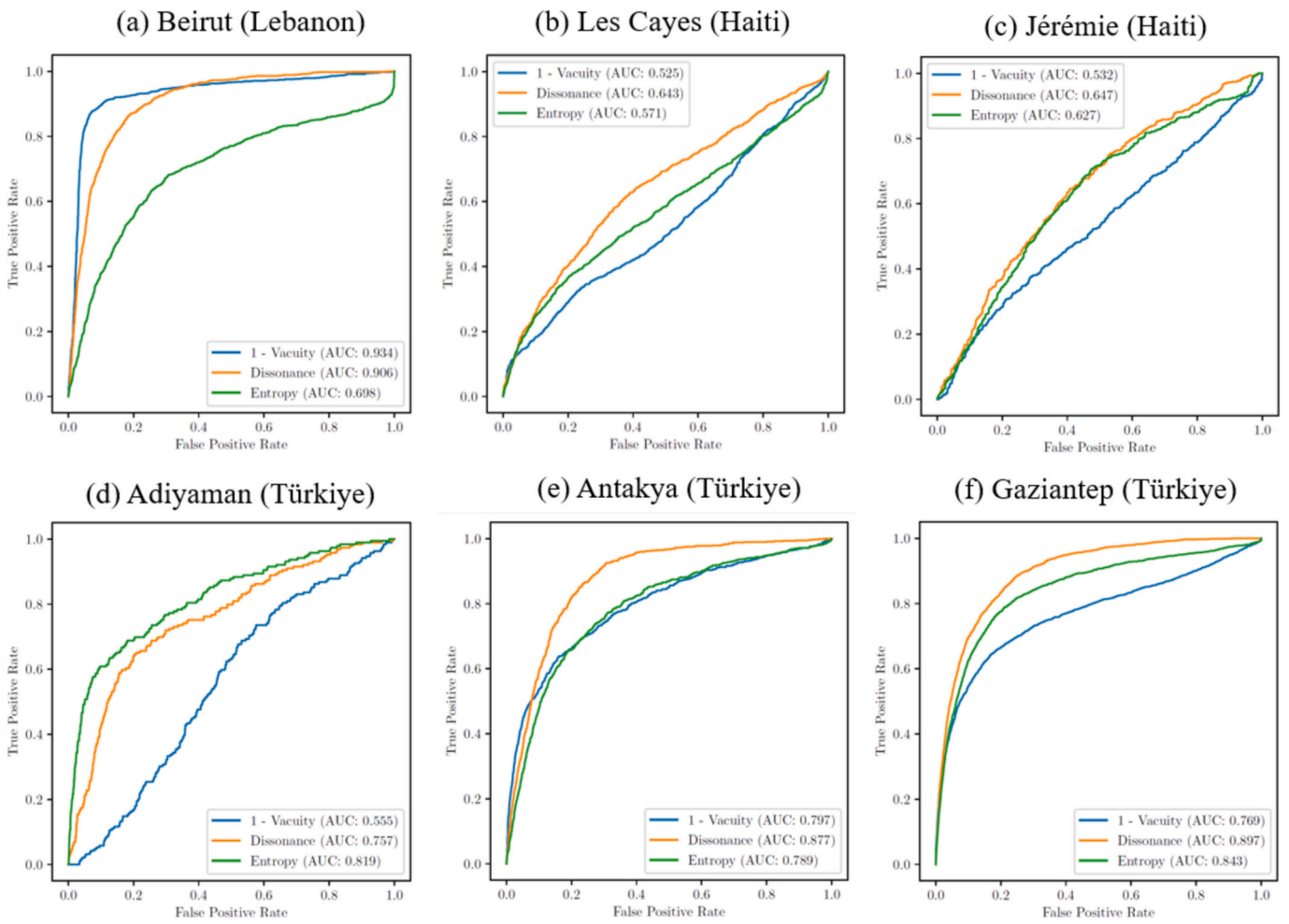

Fig. 13. Receiver operating characteristic (ROC) curves as an indication of whether each of the uncertainty measures can be used to indicate confidence in the model's predictions. Entropy is provided as a reference measure traditionally used, and it is compared with "vacuity" and "dissonance" as proposed in this work. Area under the curve (AUC) values are provided as a measure of separability (how well the model is able to separate classes).

Table 2
Performance comparison of the proposed uncertainty-aware graph analysis scheme with state of the art algorithm proposed in Chen and Zhang [48] for Beirut and Haiti case studies. Accuracy and execution time are reported when running the algorithms on a laptop with 4 CPUs and 13GB RAM, i.e. a typical platform for USAR units on the ground.

Case study	Accuracy [%]		Execution time [min]	
	Uncertainty-aware graph analysis	Bayesian NN	Uncertainty-aware graph analysis	Bayesian NN
Beirut	75 ± 2.8	76 ± 2.4	28	123
Haiti	74 ± 3.3	75.8 ± 2.6	29	144

data, and which could be refined as on-site information becomes available. The second methodology aimed to assess the uncertainty level associated with the results produced by a modified graph convolutional network performing damage classification tasks. The experimental findings illustrate the potential of these frameworks to attain the desired operational outcomes while addressing robustness and scalability. These methods represent an enhancement with respect to the current USAR post-disaster operational scenarios, from a faster means of initial rapid assessment to computational efficiency over traditional analysis methods for similar quantities of data. It is expected that these strategies could be implemented in the future in operational pipelines to streamline the use of remote sensing data analysis in emergency response.

Future works will further investigate and mitigate the effect of the non-idealities in the datasets (e.g. unbalanced datasets) that typically affect the remote sensing and other datasets typically available during search and rescue operations. There is also work to be done in further considering the dynamic aspect of the situation – immediately after an earthquake presents one damage scenario, but this may be different by the time a satellite passes, or after a further aftershock or after SAR or demolition and clearing works. The flexibility of considering graph-based methods would provide a good starting point for these investigations.

Funding

This work was supported by a grant from IStructE Earthquake Engineering Field Investigation Team (EEFIT) and Newnham College Senior Members fund. The authors are also funded by the Isaac Newton Trust, UK Engineering and Physical Sciences Research Council (EPSRC) [grant number EP/T517847/1], UKRI Centre for Doctoral Training in Application of Artificial Intelligence to the study of Environmental Risks [grant number EP/S022961/1], Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA) funded by the Research Council of Norway [RCN Grant no. 237906], and the Visual Intelligence Centre for Research-based innovation funded by the Research Council of Norway [RCN Grant no. 309439].

CRediT authorship contribution statement

Sivasakthy Selvakumaran: Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Iain Rolland:** Validation, Methodology, Formal analysis, Data curation. **Luke Cullen:** Validation, Methodology, Formal analysis, Data curation. **Rob Davis:** Writing – review & editing, Project administration. **Joshua Macabuag:** Writing – review & editing, Project administration, Conceptualization. **Charbel Abou Chakra:** Writing – review & editing, Resources, Project administration, Data curation. **Nanor Karageozian:** Writing – review & editing, Resources, Project administration, Data curation. **Amir Gilani:** Writing – review & editing, Resources, Project administration, Data curation. **Christian Geiß:** Writing – review & editing, Resources. **Miguel Bravo-Haro:** Writing – review & editing, Data curation. **Andrea Marinoni:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgements

Damage assessment information collected on the ground following the Beirut explosion was provided by UN-Habitat, with permission from the Municipality of Beirut. Damage assessment information collected on ground following the Haiti earthquake was provided by Miyamoto International, with permission from the government of Haiti. TerraSAR-X and high resolution optical images over Beirut were provided via the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt). High resolution optical images over Haiti and Türkiye before and after the event were made available via Maxar following the earthquake.

The authors would like to thank the other members of SARAID's engineering team, specifically Mark Scorer, Suzie Cooper, Giles Waley and Filip Kirazov who were also part of SARAID engineer deployments supporting damage assessment and coordination during the events studied.

References

[1] Davis I, Alexander D. Recovery from disaster. 1st ed. Routledge; 2015. https://doi.org/10.4324/9781315679808.
[2] Dalla Mura M, Prasad S, Pacifici F, Gamba P, Chanussot J, Benediktsson JA. Challenges and opportunities of multimodality and data fusion in remote sensing. Proc IEEE 2015;103(9):1585–601. https://doi.org/10.1109/JPROC.2015.2462751.
[3] Moya L, Geiß C, Hashimoto M, Mas E, Koshimura S, Strunz G. Disaster intensity-based selection of training samples for remote sensing building damage classification. IEEE Trans Geosci Remote Sens 2021;59(10):8288–304. https://doi.org/10.1109/TGRS.2020.3046004.
[4] Persello C, Wegner JD, Hänsch R, Tuia D, Ghamisi P, Koeva M, et al. Deep learning and earth observation to support the sustainable development goals: current approaches, open challenges, and future opportunities. IEEE Geosci Remote Sens Mag 2022;10(2):172–200. June 2022, https://doi.org/10.1109/MGRS.2021.3136100.
[5] Dong L, Shan J. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. ISPRS J Photogramm Remote Sens 2013;84:85–99. https://doi.org/10.1016/j.isprsjprs.2013.06.011.
[6] Geiß C, Taubenböck H. Remote sensing contributing to assess earthquake risk: from a literature review towards a roadmap. Nat Hazards 2013;68:7–48. https://doi.org/10.1007/s11069-012-0322-2.
[7] Gupta R, Hosfelt R, Sajeev S, Patel N, Goodman B, Doshi J, et al. xBD: a dataset for assessing building damage from satellite imagery; 2019. Arxiv; 2019 [online] Available at: https://arxiv.org/abs/1911.09296.
[8] Gokon H, Post J, Stein E, Martinis S, Twele A, Mück M, et al. A method for detecting buildings destroyed by the 2011 Tohoku earthquake and tsunami using multitemporal TerraSAR-X data. IEEE Geosci Remote Sens Lett 2015;12(6):1277–81. https://doi.org/10.1109/LGRS.2015.2392792.
[9] Plank S. Rapid damage assessment by means of multi-temporal SAR-A comprehensive review and outlook to Sentinel-1. Remote Sens (Basel) 2014;6(6). https://doi.org/10.3390/rs6064870.
[10] Kaufman YJ, Ichoku C, Giglio L, Korontzi S, Chu DA, Hao WM, et al. Fire and smoke observed from the earth observing system MODIS instrument–products, validation, and operational use. Int J Remote Sens 2003;24(8):1765–81. https://doi.org/10.1080/01431160210144741.
[11] Ez-zahouani B, Teodoro A, El Kharki O, Jianhua L, Kotaridis I, Yuan X, et al. Remote sensing imagery segmentation in object-based analysis: a review of methods, optimization, and quality evaluation over the past 20 years. Remote Sens Appl Soc Environ 2023;32:101031. https://doi.org/10.1016/j.rsase.2023.101031.
[12] Chlaily S, Dalla Mura M, Chanussot J, Jutten C, Gamba P, Marinoni A. Capacity and limits of multimodal remote sensing: theoretical aspects and automatic information theory-based image selection. IEEE Trans Geosci Remote Sens 2021;59(7):5598–618. https://doi.org/10.1109/TGRS.2020.3014138.
[13] Lahat D, Adali T, Jutten C. Multimodal data fusion: an overview of methods, challenges, and prospects. Proc IEEE 2015;103(9):1449–77. https://doi.org/10.1109/JPROC.2015.2460697.
[14] Gomez-Chova L, Tuia D, Moser G, Camps-Valls G. Multimodal classification of remote sensing images: a review and future directions. Proc IEEE 2015;103(9):1560–84. https://doi.org/10.1109/JPROC.2015.2449668.
[15] Brunner D, Lemoine G, Bruzzone L. Earthquake damage assessment of buildings using VHR optical and SAR imagery. IEEE Trans Geosci Remote Sens 2010;48(5):2403–20. https://doi.org/10.1109/TGRS.2009.2038274.
[16] Geiß C, Aravena Pelizari P, Marconcini M, Sengara W, Edwards M, Lakes T, et al. Estimation of seismic building structural types using multi-sensor remote sensing and machine learning techniques. ISPRS J Photogramm Remote Sens 2014;104:175–88. https://doi.org/10.1016/j.isprsjprs.2014.07.016.
[17] Adriano B, Yokoya N, Xia J, Miura H, Liu W, Matsuoka M, et al. Learning from multimodal and multitemporal earth observation data for building damage mapping. ISPRS J Photogramm Remote Sens 2021;175:132–43. https://doi.org/10.1016/j.isprsjprs.2021.02.016.
[18] Macabuag J. How can structural engineers contribute towards disaster mitigation? Struct Eng 2013;91(6). Available at: https://www.istructe.org/journal/volumes/volume-91-(2013)/issue-6/feature-how-can-structural-engineers-contribute-to/.
[19] Taubenböck H, Goseberg N, Setiadi N, Lämmel G, Moder F, Oczipka M, et al. "Last-mile" preparation for a potential disaster - interdisciplinary approach towards tsunami early warning and an evacuation information system for the coastal city of Padang, Indonesia. Nat Hazards Earth Syst Sci 2009;9(4):1509–28. https://doi.org/10.5194/nhess-9-1509-2009.
[20] Voigt S, Kemper T, Riedlinger T, Kiefl R, Scholte K, Mehl H. Satellite image analysis for disaster and crisis-management support. IEEE Trans Geosci Remote Sens 2007;45(6):1520–8. https://doi.org/10.1109/TGRS.2007.895830.
[21] United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA). Volume II: preparedness and response, INSARAG guidelines. 2020.
[22] Meresse P, Fabrizio G, Mori M, Russo M, Annecchini A, Fanfani A, et al. Manual of ASR1 – wide area assessment process [online] Available at: https://www.beliceproject.eu/wp-content/uploads/2021/04/BELICE-MANUAL-EN.pdf; 2021.
[23] Davis R. The search and rescue assistance in disasters (SARAID) response to the Beirut explosion68. International Fire Fighter; 2020. p. 22–4.
[24] Macabuag J, Altheim C, Thorvaldsdottir S, Perks D. Damage assessments by international engineers following the Albania earthquake of November 2019. Int J Disaster Risk Reduct 2019;72:102822. https://doi.org/10.1016/j.ijdrr.2022.102822.
[25] World Bank. Global rapid post-disaster damage estimation (GRADE) report: February 6, 2023 Kahramanmaraş earthquakes - Türkiye report (English). Report accessed online December 20th 2023: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099022723021250141/p178843 0aeb62f08009b2302bd4074030fb; 2023.
[26] World Bank Group, European Union, & the United Nations. Beirut rapid damage and needs assessment. Report accessed online December 20th 2023: https://documents1.worldbank.org/curated/en/650091598854062180/Beirut-Rapid-Damage-and-Needs-Assessment.pdf; 2020.
[27] UNOCHA. The humanitarian response to the Beirut port explosions: Lebanon flash appeal 2020 – end report (30 April 2021). Report accessed online December 20th 2023, https://www.humanitarianresponse.info/sites/www.humanitarianresponse.info/files/documents/files/end_of_fa_report.pdf; 2021.
[28] Agapiou A. Damage proxy map of the Beirut explosion on 4th of august 2020 as observed from the copernicus sensors. Sensors (Switzerland) 2020;20(21):1–21. https://doi.org/10.3390/s20216382.
[29] Pilger C, Gaebler P, Hupe P, Kalia AC, Schneider FM, Steinberg A, et al. Yield estimation of the 2020 Beirut explosion using open access waveform and remote sensing data. Nat Sci Rep 2021;11(1):1–14. https://doi.org/10.1038/s41598-021-93690-y.
[30] Municipality of Beirut & UN-Habitat. Beirut port explosions response: Beirut municipality rapid building-level damage assessment. Available online, https://unhabitat.org/sites/default/files/2020/10/municipality_of_beirut_-_beirut_explosion_rapid_assessment_report.pdf; 2020.

[31] Municipality of Bourj Hammoud and UN-Habitat. Beirut port explosions response: Bourj Hammoud municipality rapid building-level damage assessment. Available online: https://unhabitat.org/sites/default/files/2020/10/municipality_of_bourj_hammoud_-_beirut_explosion_rapid_assessment_report.pdf; 2020.
[32] UNOSAT (United Nations Satellite Centre). EQ20210814HTI: 14 August 2021 M7.2 Haiti earthquake. Preliminary satellite-based comprehensive damage assessment report. Report accessed online November 30th 2022, https://data.humdata.org/dataset/14-august-2021-m7-2-haiti-earthquake-preliminary-satellite-based-comprehensive-damage-asse; 27 August 2021.
[33] NASA Jet Propulsion Laboratory. Satellite data shows ground motion from August 2021 Haiti earthquake. August 17th 2021. Available online: https://www.jpl.nasa.gov/images/pia23697-satellite-data-shows-ground-motion-from-august-2021-haiti-earthquake; 2021.
[34] US Geological Survey. M7.8 and M7.5 Kahramanmaraş earthquake sequence near Nurdağı, Turkey (Türkiye). Reston, VA, USA: US Geological Survey; 2023. Report accessed online March 3rd, 2023: https://www.usgs.gov/news/featured-story/m78-and-m75-kahramanmaras-earthquake-sequence-near-nurdagi-turkey-turkiye.
[35] Whitworth MRZ, Giardina G, Penney C, Di Sarno L, Adams K, Kijewski-Correa T, et al. Lessons for remote post-earthquake reconnaissance from the 14 August 2021 Haiti earthquake. Front Built Environ 2022;8(April):1–16. https://doi.org/10.3389/fbuil.2022.873212.
[36] Gunasekera R, Ishizawa Escudero OA, Daniell JE. Global rapid post-disaster damage estimation (GRADE) report: February 6, 2023 Kahramanmaraş earthquakes - Türkiye report. DC, USA: World Bank Group Washington; 2023 [online] Available at: http://documents.worldbank.org/curated/en/099022723021250141/P1788430aeb62f08009b2302bd4074030fb (accessed 09/03/2023).
[37] Dong X, Thanou D, Toni L, Bronstein M, Frossard P. Graph signal processing for machine learning: a review and new perspectives. IEEE Signal Process Mag 2020; 37:117–27. https://doi.org/10.1109/MSP.2020.3014591.
[38] Hamilton WL. Graph representation learning Hamilton. Synth Lect Artif Intell Mach Learn 2020;14(3):1–159. ISSN 19394616.
[39] DiMaio F, Shavlik J. Belief propagation in large, highly connected graphs for 3D part-based object recognition. In: Sixth international conference on data mining (ICDM'06), Hong Kong, China; 2006. p. 845–50. https://doi.org/10.1109/ICDM.2006.26.
[40] Yamaguchi Y, Faloutsos C, Kitagawa H. CAMLP: confidence-aware modulated label propagation. In: 16th SIAM international conference on data mining 2016 (SDM 2016), Miami, Florida, USA, May 57, 2016; 2016. p. 513–21. https://doi.org/10.1137/1.9781611974348.58.
[41] Zhao X, Chen F, Hu S, Cho JH. Uncertainty aware semi-supervised learning on graph data. In: Conference on neural information processing systems (NeurIPS 2020), virtual, December 6-12, 2020. 33; 2020. p. 12827–36.
[42] Eswaran D, Günnemann S, Faloutsos C, Makhija D, Kumar M. ZooBP: belief propagation for heterogeneous networks. Proc VLDB Endow 2017;10(5):625–36. https://doi.org/10.14778/3055540.3055554.
[43] Rolland I, Marinoni A, Selvakumaran S. Uncertainty-aware graph-based multimodal remote sensing detection of out-of-distribution samples. In: CDCEO 2021: 1st workshop on complex data challenges in earth observation, November 1, 2021. QLD, Australia: Virtual Event; 2021. https://ceur-ws.org/Vol-3052/short5.pdf.
[44] Jøsang A, Cho J-H, Chen F. Uncertainty characteristics of subjective opinions. In: 21st international conference on information fusion (FUSION 2018), Cambridge, U.K., July 10-13, 2018; 2018. p. 1998–2005.
[45] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks; 2016. arXiv preprint; 2016. https://arxiv.org/abs/1609.02907.
[46] Jøsang A. Subjective logic - a formalism for reasoning under uncertainty. In: Artificial intelligence: foundations, theory, and algorithms. Springer; 2016. https://doi.org/10.1007/978-3-319-42337-1.
[47] Marinoni A, Chlaily S, Khachatrian E, Eltoft T, Selvakumaran S, Girolami M, et al. Enhancing ensemble learning and transfer learning in multimodal data analysis by adaptive dimensionality reduction; 2021. Arxiv; 2021 [online] Available at: https://arxiv.org/abs/2105.03682v1.
[48] Chen W, Zhang L. Predicting building damages in mega-disasters under uncertainty: an improved Bayesian network learning approach. Sustain Cities Soc 2021;66:102689. https://doi.org/10.1016/j.scs.2020.102689.