ORIGINAL RESEARCH

# From Research to Applications: What Can We Extract with Social Media Sensing?

Aris Bozas[1] · Stelios Andreadis[1] · Despoina Chatzakou[1] · Spyridon Symeonidis[1] · Ourania Theodosiadou[1] · Pantelis Kyriakidis[1] · Alexandros Kokkalas[1] · Evangelos A. Stathopoulos[1] · Sotiris Diplaris[1] · Theodora Tsikrika[1] · Ilias Gialampoukidis[1] · Stefanos Vrochidis[1] · Ioannis Kompatsiaris[1]

Received: 28 July 2023 / Accepted: 13 February 2024

**Abstract**
With the constant growth of social media in our daily lives, a huge amount of information is generated online by multiple social networks. However, what can we actually extract with the science of social media sensing? It is a very challenging task to mine meaningful data out of this vast crowdsourcing volume, which also rapidly changes or ends up being misleading. The scope of this paper is to present different approaches that overcome these challenges and utilize social media information from various sources. This work illustrates applications that: improve the performance of architectural design; preserve the cultural heritage; enhance citizen security; provide early detection for disasters; and discover creeping crisis events. A large variety of analyses are presented, including, among other, disaster or crime event detection, user identity linkage, relevance classification, and community detection techniques. The evaluation of the presented methods is also given in this article, proving that they can be practical and valuable in many applications.

**Keywords** Social media sensing · Architectural design · Citizen security · Disaster management · Creeping crisis

## Introduction

Used by millions of people worldwide, social media are Internet-based channels that allow users to interact opportunistically and present themselves selectively, either in real time or asynchronously, with a wide and narrow audience,

This article is part of the topical collection "Signal Processing and Multimedia Applications" guest edited by Andrew Sung and Simone Santini.

✉ Aris Bozas
arbozas@iti.gr

Stelios Andreadis
andreadisst@iti.gr

Despoina Chatzakou
dchatzakou@iti.gr

Spyridon Symeonidis
spyridons@iti.gr

Ourania Theodosiadou
raniatheo@iti.gr

Pantelis Kyriakidis
pantelisk@iti.gr

Alexandros Kokkalas
akokkalas@iti.gr

Evangelos A. Stathopoulos
estathop@iti.gr

Sotiris Diplaris
diplaris@iti.gr

Theodora Tsikrika
theodora.tsikrika@iti.gr

Ilias Gialampoukidis
heliasgj@iti.gr

Stefanos Vrochidis
stefanos@iti.gr

Ioannis Kompatsiaris
ikom@iti.gr

[1] Information Technologies Institute, Centre for Research and Technology Hellas, 6th Km. Charilaou - Thermi Road, 57001 Thessaloniki, Greece

Published online: 24 April 2024

which derives value from user-generated content and the perception of interacting with others [1]. Social networks are a highly dynamic source of data that reflects events and the evolution of community focus and user interests. The penetration of smartphones and mobile devices into people's daily lives provides a continuous real-time and location-based user feedback that can be used for various applications. Due to its increasing popularity and influence, a large number of studies have focused on investigating the use of these platforms by the general public in their daily lives to convert individual infrequent, but collectively frequent media into meaningful topics, events, points of interest, emotional states, and social connections.

In the last decade, social media platforms have been able to provide us with huge amounts of raw data in the form of text, images, video and audio, and are rich in social context, as social connections and relationships can be extracted using social graphs and can been used for a variety of applications on various popular topics of interest; to name a few, crisis events [2], natural disasters [3, 4], migration events [5, 6], tourism issues [7] and security [8].

However, social media have many challenges that need to be surpassed in order for their content to be used in applications in an effective way. For example, most of the texts are short, conversations most of the time contain errors and provide no context. In addition, lots of posts on social networks constitute noise and are therefore unsuitable for exploitation, as they are either spam, fake news or propaganda. Additionally, the existence of a huge volume of constantly changing data, as they are dynamically generated in real time with rapid updates and most of the time by different communities and different applications, makes it difficult to monitor all this information.

This paper presents techniques to address these challenges and demonstrates applications for extracting complementary knowledge that can be used in various domains by leveraging crowdsourced information from social media. More specifically, this work suggests applications in the architectural design domain that provides inspiration to enhance and make the work of professionals more efficient by automatically retrieving and converting web-based data that can be imported into architectural design projects from YouTube, Flickr, and Google reviews. In addition, two cultural applications are presented. The first one aims to preserve intangible cultural heritage and the second one to preserve the cultural heritage of refugees with textual content from Twitter and visual content from YouTube. This paper also presents methods for crisis event detection, crime detection and prevention, as well as a method for identifying accounts that possibly belong to the same individual. Furthermore, a methodology is proposed to detect events for disaster, water quality and air quality incidents, by acquiring and processing tweets in near real-time. Also, a relevance estimation methodology is presented that separates relevant posts from irrelevant posts for the disaster domain (e.g., floods, fires, earthquakes). Finally, a framework that makes use of various crowdsourced information from citizens about water-related alerts (e.g., debris, oil spills, algal blooms) is described.

The rest of the work is organised as follows. In the section "Related Work" we list related work that exploits the social media in the areas of (i) architectural design, (ii) cultural applications, (iii) crime prediction, detection and prevention, (iv) disaster management, and (v) water management. In the section "Overview of Our Approaches in Social Sensing" various applications and methodologies related to the above-mentioned domains are presented and the reader is provided with all the necessary background. Next, the results and the evaluation of these applications and methodologies are described in the section "Representative Results". The section "Conclusions and Future Research Directions" concludes the article and suggests future directions for the work presented.

## Related Work

Social media are a relatively new tool for the people to express their thoughts, sentiments or report disasters and crime activity and much more. The explosive growth of them resulted in the daily generation of a vast amount of information that, if collected and processed efficiently, is a strong asset for applications and can improve our lives in essentially every aspect. Below, several studies from several domains that utilized social data are presented.

In the architecture design domain, social media have been already investigated as sources of inspiration and supplementary knowledge. It is recognized that, if processed in an appropriate manner, they have the potential to provide valuable insights and improve the professionals' workflows. In the work of [9], Instagram posts from multiple users have been leveraged to obtain feedback about site-scale landscape design issues in the Seattle Freeway Park. Specifically, the goal was to find answers to questions about they way it is used, considering also the temporal and climate factors, and the people's emotional attitude towards it. In a similar effort, [10] used content from the Twitter platform for post-occupancy evaluation of the High Line in New York City. The data have been pre-processed using text mining techniques and analyzed in three dimensions (content, network and spatial sentiment analysis). The findings of the study show the capabilities of social media analytics as a tool that can be used in place of conventional methods (e.g., surveys) for public space design evaluation and rapid elicitation of user feedback.

The value of social media in urban design has also been explored in the architecture education domain. Valls et al.

[11] acquired data from Twitter and Flickr with respect to Gillet Square in London, an award-winning public space, and analyzed them to extract semantic, temporal and spatial patterns that highlight the strategic requirements of an urban design project. The analysis helps teaching staff define the project program of a course considering aspects that were typically overlooked due to the lack of relevant data, and equips students with visually engaging education materials that support their creative design process.

Citizens' security both in offline and online world is of major importance and can be significantly enhanced by the development and use of modern methodologies and tools that are often based on Artificial Intelligence. Nowadays, social media have become an integral part of the life of more than 4.7B people around the globe. Due to their intense use by a large portion of people around the world, they have turned into a very powerful instrument for shaping public opinion and spreading all kinds of not always harmless content. The timely and valid analysis of the ever-increasing amount of data generated on a daily basis in social media can enable the extraction of useful information and conclusions about key issues, or even the prevention of potential criminal behaviour in the online as well as the offline world. For instance, in a study of a violent incident (shooting of four police officers) in the Seattle-Tacoma, Washington, it was demonstrated that the majority of the messages posted on Twitter, regarding such incident, contained useful information [12].

As shown with the Seattle-Tacoma example, the collection and analysis of content produced in social media platforms can play a vital role in almost real-time incident detection. To this end, in the recent decade, various approaches for Event Detection have been presented. From an extraction point of view, the goal of Event Detection is to extract *who, did-what, where* and other argument types from the text, such as involved entities. This paradigm has been studied extensively in a well structured document collection [13], either using handcrafted features [14–16] or more recently, deep learning architectures such as Convolutional Neural Networks (CNNs) [17, 18], Recurrent Neural Networks (RNNs) [19], Graphs [20, 21], and Attention mechanisms [22].

While these algorithms perform well in this information-rich scenario, there is little effort in algorithms that categorise the event-related informativeness of texts in noise-filled, unstructured social media data. A well-known and widely used dataset in this context is CrisisLexT26 [23], which is publicly available and contains Twitter posts from 26 different crisis events; posts are labelled based on the relatedness to each crisis event, on the information source (e.g., governments, NGOs) and type (e.g., caution and advice, infrastructure damage). While this dataset has been investigated using CNNs (e.g., [24–27]), limited emphasis was given to dealing with the noisy nature of this data (short text, informal language, grammatical problems, unordered summaries, etc.) [28].

In addition, social media have been exploited to commit delinquent behaviour or even to recruit terrorists and criminals online [29]. Aiming to early detect or even prevent possible criminal behaviours based on the collected online content, Law Enforcement Agencies could exploit the outcomes derived from the analysis of the temporal evolution of various terrorism or crime-related indicators extracted from online activity. Towards this direction, the implementation of change point analysis in time series data of interest could be of great importance. This analysis enables the detection of significant changes that take place over time which may possibly signify the occurrence of events of interest that attention should be paid to.

Concerning the use of change point detection (CPD) in content obtained from social media sources, it has mainly been applied to time series related to the frequencies of Twitter posts so as to estimate the time locations of statistical significant changes in the evolution of the series and explore the connection between the identified time points and the occurrence of events (e.g., [30–32]). Focusing on more infringing content (e.g., terrorism-related), CPD has been applied to the Noordin Top terrorist network data [33] to detect significant changes in the evolution of its structure. In addition, a multiple CPD framework has been applied in [34] concerning multivariate time series derived by the frequencies of terrorism events across twelve global regions.

Moreover, users with delinquent online behaviour (which is often reflected in the offline world) in an effort to multiply the spread of their thoughts, ideas, and viewpoints (e.g., to recruit to a greater extent), tend to create multiple accounts to bypass the combating measures enforced by social media platforms and thus retain their online identity even if some of their accounts are suspended. In this direction, *identity resolution* (or *identity linkage*) has emerged, aiming to uncover links among the unprecedented high number of online user accounts [35]. Identity resolution has also been used by Law Enforcement as a way to uncover previously unknown connections between actors [36]. Thus far, numerous studies have focused on identifying potentially linked accounts, mainly *across* social media platforms, as for instance, between Twitter and Foursquare [37, 38], Instagram [37], or Facebook [39, 40]. With the focus being more on forums, websites, and online discussion communities rather than on social media platforms, identity resolution has also been explored *within* a single social network. For instance, identity linkage has been researched on popular online news sites, such as *The Guardian* and the *SPIEGEL ONLINE*, to assist their providers detect manipulations of public opinion [41]. Moreover, an Irish forum has been studied [42] to first unmask authors identities and then detect matching aliases. Finally, in relation to social media platforms, Twitter has been considered [43, 44] to this end.

Furthermore, social media has an important role in cultural applications as it has become an essential component of everyday life. Social media platforms provide strong public venues with vast audiences for people to freely discuss their thoughts and views on a variety of issues in the world. More specifically, a study concerning refugee crisis social sensing [45], used Twitter data of 35,000 relevant posts from five different languages to analyze public sentiment on the matter. Using supervised machine learning algorithms, the study provides a binomial classification of positive and negative comments for posts supporting and posts that are against the refugees. Another area of application is applied for mapping historic urban landscape values [46]. In this study, data of photos shared on Flickr were used in order to analyze the perceived landscape character that determines heritage for each person. The results show the difference of preferences between the locals and the tourists as well as the difference of experts' and users' definition of heritage suggesting that a strict and closed categorization fails to include the various and diverse aspects of heritage. Another study [47] uses social media and crowdsourcing, to extract local knowledge and reveal intangible cultural heritage for urban planning. The data were collected through PPGIS,[1] Instagram and Facebook groups. The case study is town Nikkilä in Finland and the aim is to reveal resident's values, place-based memories and experiences comparing non-professional knowledge with expert knowledge.

Additionally, social media platforms can play an essential role in the emergency response by providing valuable information for early warning tools and produce an accurate picture of the situational awareness during a disaster. Focusing on disaster management, Sherchan et al. [48] designed an automated tool called Australian Crisis Tracker (ACT) provided with Web interface for users to explore events and effortlessly discover information in order to accelerate response and recovery during a natural disaster. ACT detects tweets referring to natural disasters by analyzing Twitter's real-time stream and they are clustered into event categories (e.g., fire, flood, storm). Then, the events are georeferenced and augmented with images from Instagram. Likewise, Chatfield and Brajawidagda [49] examined the utility of Twitter as an alternative early tsunami warning system for the region of Indonesia. The content analysis of tweets from agencies related to tsunami warning shows that Twitter can provide the public with a time lead of approximately 7 min for the residents of vulnerable areas to evacuate. Moreover, Aulov et al. [50] make use of geolocated data from Twitter and Flickr in order to forecast the trajectory of oil spills by converting the social media posts into geophysical data (e.g., latitude, longitude, timestamp) and utilizing them as boundary conditions in the General NOAA Oil Modeling Environment (GNOME) for oil predictions to obtain improved estimates of the model parameters.

On top of that, social media can provide assistance in evaluation and monitoring of the water quality as surface freshwater is a finite resource necessary to the survival of mankind and drinking safe water is of the utmost importance for the consumer health for avoiding severe illnesses and diseases. The work of Lambert and Bir [51] evaluates water quality by using sentiment analysis on the social media posts of Twitter, Instagram, blogs, forums and websites related to tap water quality by using sentiment analysis. In another work, Zheng et al. [52] propose a framework for surface water quality monitoring through a We-chat based application that receives water quality reports from volunteers. Moreover, Wickstrom and Specht [53] aims to identify high-influence Twitter participants of a water-quality discussion about a water ban in Toledo in order to detect emerging water issues from the influential accounts and to discover the authoritative users in their geographic regions.

## Overview of Our Approaches in Social Sensing

In this section, we describe several methodologies and techniques that incorporate social media data. The upcoming the section "Architecture Design" concerns two cases where social data are used for the domain of architecture. In the section "Strengthening Citizens' Security" three approaches are introduced in the domain of security for crisis and crime detection and prevention, while two cultural applications are suggested for the protection of cultural heritage in the section "Cultural Applications". Moreover, the section "Disaster Management" illustrates techniques for estimating the context relevance of tweets in certain topics and for detecting disaster events and creeping crises. Lastly, the section "Water Management" refers to methodologies for the identification of water quality incidents.

### Architecture Design

Social media, if exploited correctly, can reasonably support architects and designers. The following sections ("Reuse and Repurpose of Multimedia for Indoor and Outdoor Designs" and "Emotionally Adaptive Design") present two cases where we gathered and processed social media content for the architecture domain.

[1] https://en.wikipedia.org/wiki/Participatory_GIS.

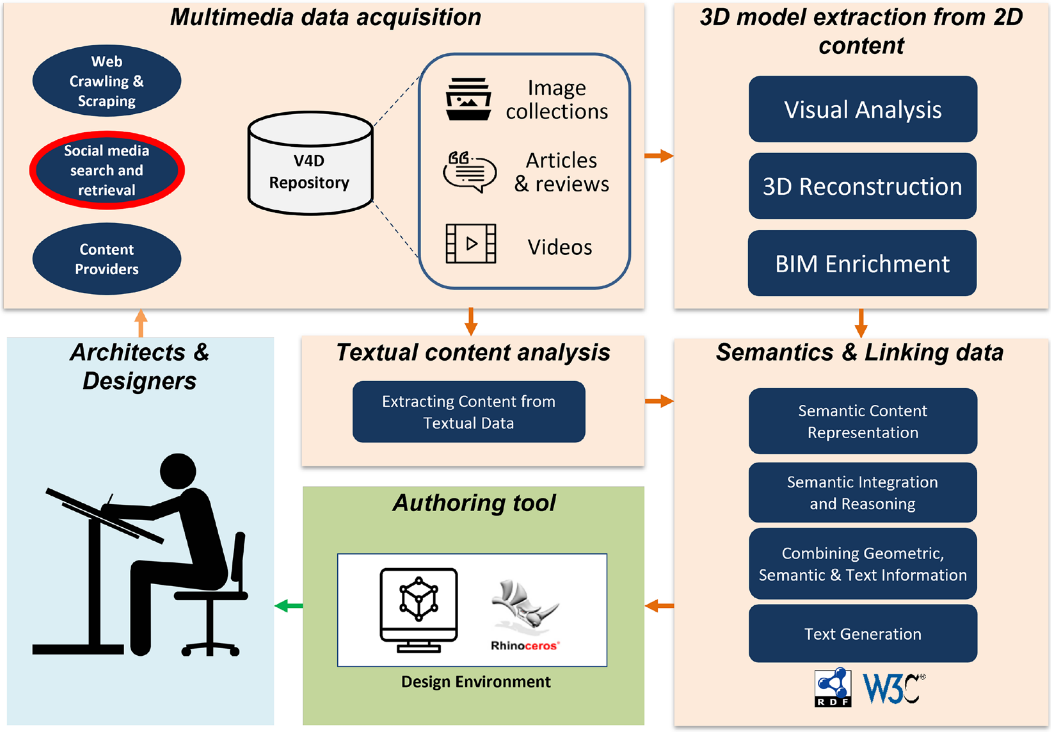


**Fig. 1** The V4Design concept

## Reuse and Repurpose of Multimedia for Indoor and Outdoor Designs

Nowadays, large amounts of new multimedia content are being produced and shared online, including archived materials that are being digitized, resulting in great troves of data of interest to the architecture design communities. The main issue is that this content in its current form is significantly under-exploited, due to the lack of appropriate solutions for its retrieval and integration into the design process.

In order to bridge this gap, we implemented the V4Design[2] system, a solution that allows for automatic content analysis and seamless transformation to assist architects (among other creative industries) in sharing content and maximizing its exploitation. V4Design provides the ability to re-use and re-purpose existing visual and textual content from content providers, as well as public Web and social media resources. To this end, it integrates and intelligently combines several state-of-the-art technologies, such as text analysis, visual analysis, 3D reconstruction and semantics. The solution for architecture is built on top of existing designing software (i.e., McNeel's commercial software, Rhino), allowing users to work directly in a well-known environment. V4Design is not just providing additional knowledge and insights to the designers; the real added value is the pipeline (shown in Fig. 1) that automatically transforms online data into semantically enriched 3D assets, which can be easily imported into an architecture design project in Rhino[3] (Fig. 2). Furthermore, in comparison with the related work, it is able to support both outdoor and indoor design scenarios. All in all, it is a rich and sophisticated tool that facilitates designers throughout the whole design procedure, from the creation of an initial concept to a final design proposal.

As the V4Design process starts with freely available online data, it is evident that social media play an important role in it. We implemented a data collection module that searches and retrieves high quality content from multiple social media in order to cater to various informational needs: (i) the Flickr image hosting website, (ii) the YouTube and Vimeo video sharing platforms, (iii) Google reviews. The visual content feeds the 3D reconstruction service, and several visual analysis modules that recognise the principle elements (e.g., objects and buildings) in imagery data (video and still image footage). Images are handled as image collections about the same object, whereas videos are split into shots before further processing. The textual content is fed to the text analysis module, which performs a preliminary analysis, extracting features, such as concepts, named entities and sentiment tags.

[2] https://v4design.eu/.

[3] https://www.rhino3d.com/.

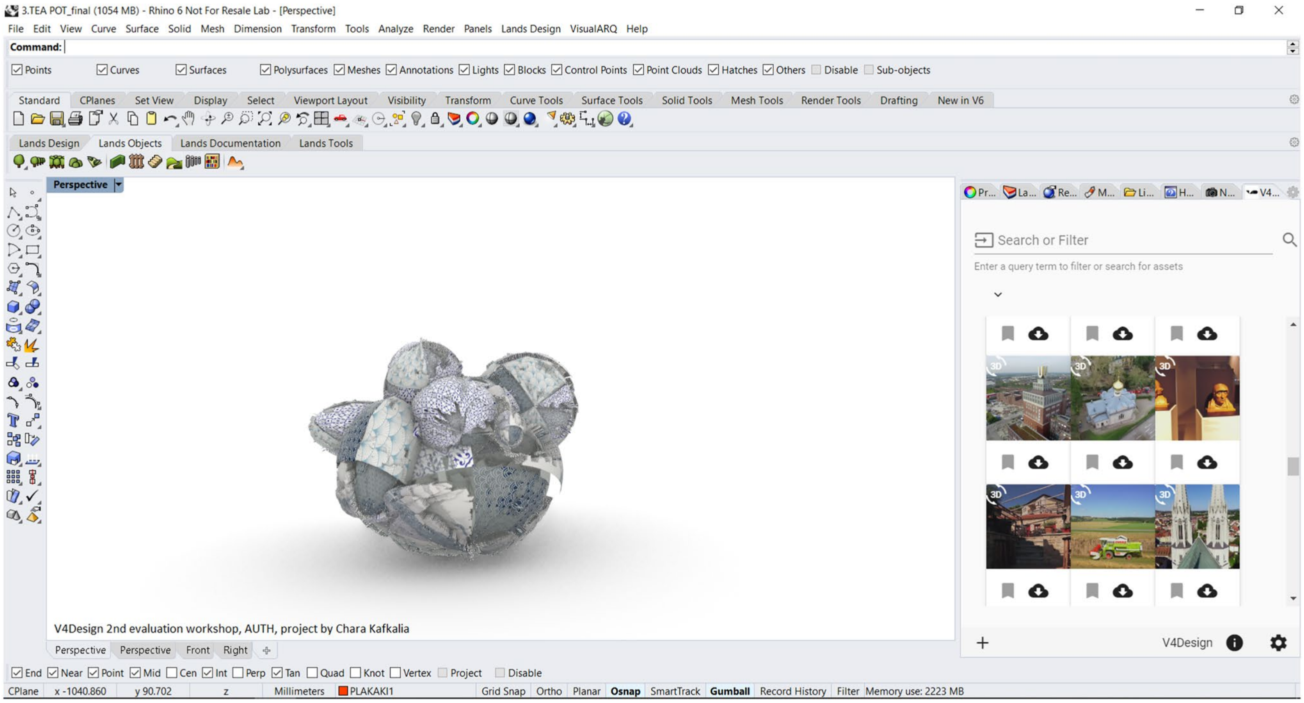

3.TEA POT_final (1054 MB) - Rhino 6 Not For Resale Lab - [Perspective]
File Edit View Curve Surface Solid Mesh Dimension Transform Tools Analyze Render Panels Lands Design VisualARQ Help
Command:
Standard CPlanes Set View Display Select Viewport Layout Visibility Transform Curve Tools Surface Tools Solid Tools Mesh Tools Render Tools Drafting New in V6
Lands Design Lands Objects Lands Documentation Lands Tools
Perspective
Search or Filter
Enter a query term to filter or search for assets
V4Design 2nd evaluation workshop, AUTH, project by Chara Kafkalia
Perspective Front Right
V4Design
CPlane x -1040.860 y 90.702 z Millimeters PLAKAKI1 Grid Snap Ortho Planar Osnap SmartTrack Gumball Record History Filter Memory use: 2223 MB

**Fig. 2** The V4Design Rhino plugin

For Flickr, we established a mechanism that takes as input a textual query, and then searches the website using the input and a filter that restricts the retrieved results to images distributed with a compatible license that allows re-use of them (e.g., Public Domain Mark, CC0). The search result is stored as an image collection related to the same object. The YouTube/Vimeo retrieval sub-module takes as input a URL and stores the linked video into the repository. The same distribution license conditions as in Flickr are applied, in other words, the video will not be stored if the license is not appropriate. Lastly, for the Google reviews, the implemented wrapper takes a textual input (e.g., of a famous landmark) and returns user reviews related to it (if they exist). In total, the V4Design main repository has been populated with 53,726 images from Flickr, 500 YouTube videos, 18 Vimeo videos and 1,273 Google reviews.

## Emotionally Adaptive Design

In another architecture design scenario, under a novel working model scheme, artists and technology experts closely collaborate to propose creative designs to address the societal challenges cities face as they grow as well as the changing requirements for functionality and emotional resonance of contemporary workplace and housing interiors. It is the scenario implemented by the Mindspaces tools[4]. Its goal is to develop tools and solutions for adaptable and inclusive places that dynamically adjust to end users' emotional, aesthetically pleasing, and social responses, producing architectural designs that are both functional and emotionally appealing. State-of-the-Art wearable electroencephalography (EEG), physiological sensing, visual analysis, and social media inputs are only a few examples of the multi-modal data and technologies that are incorporated for the quick evaluation of natural user reactions to the MindSpaces installations and artistic adaption of the designs accordingly. The end user tools are built on top of Rhino and Unity[5]). The high level concept of the developed system is illustrated in Fig. 3.

To fulfil the social media content needs, we implemented a module that obtains freely available textual and visual data from Twitter and selected websites that contain user comments. Specifically, the target data consists of images of architectural interest, tweets related to urban and interior design, as well as user posts and user comments from thematic Web resources. These multimedia items fueled the project's textual and visual analysis modules. For the former, more details regarding the extracted knowledge of the text analysis pipeline exist in [54].

For retrieving Twitter posts, a search mechanism was set up that was querying the Twitter platform using textual keywords. To compile the required input (i.e., a group

[4] https://mindspaces.eu/.

[5] https://unity.com/developer-tools.

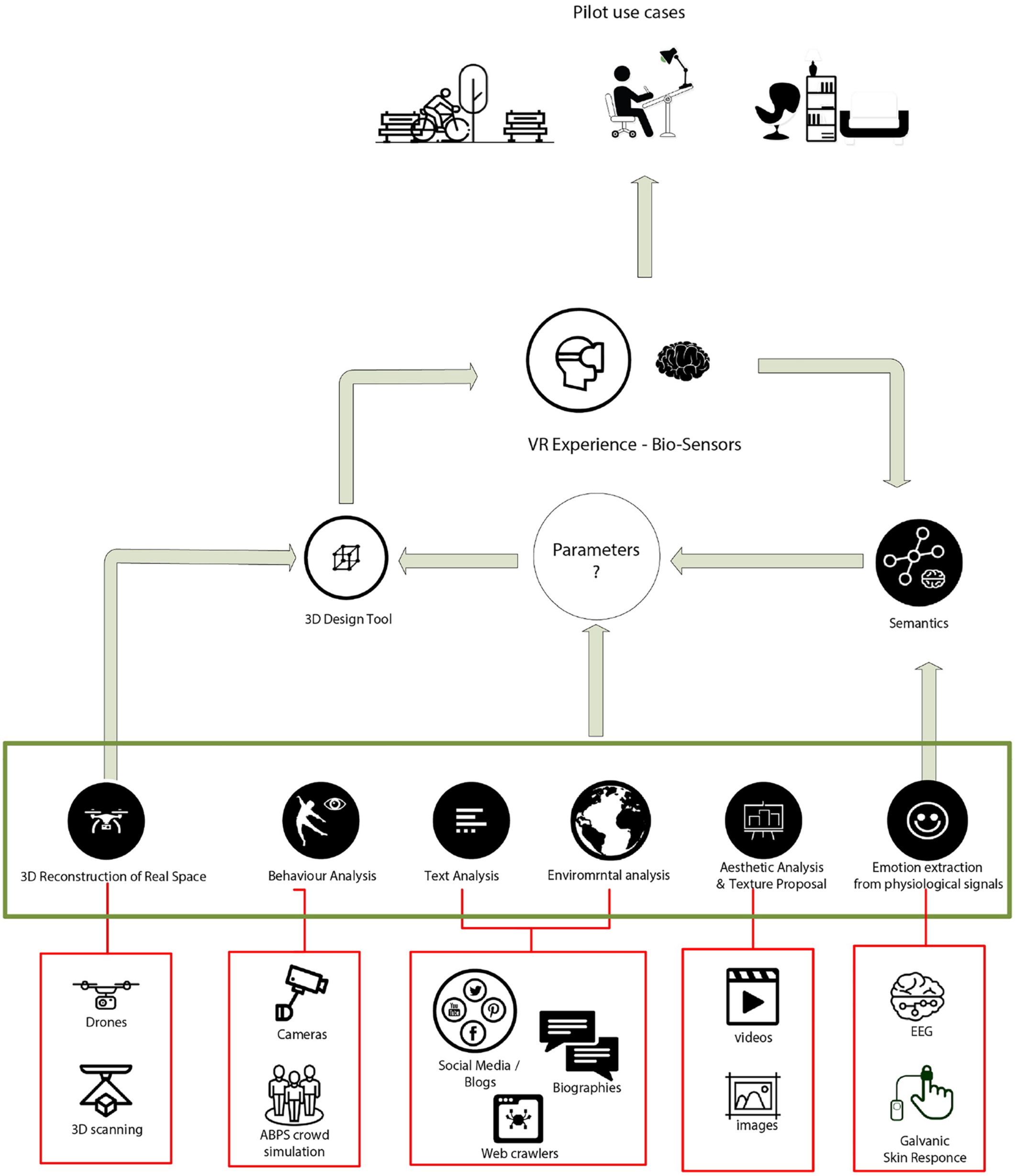


**Fig. 3** The Mindspaces concept

of keywords related to art and culture in urban environments), a survey was conducted by the platform's users. The outcome comprises the following groups/topics: Planning of the City, Art, Social Impact, Sensations, Idioms, People, Coexistence, Insulation of Community, Integration of Community, Accessibility of the Community, and Insecurity of the Community. The search module has been running for five weeks, iteratively querying each keyword group in English, Spanish and Catalan languages, resulting in collections of 155,000, 190,000 and 55,000 tweets respectively.

Concerning the online sources of a particular theme, Web crawling and scraping has been applied. There were numerous options for finding thematic websites of interest; nevertheless, only a few were selected based on well-defined structures. Details about the chosen websites are presented in Table 1. Deep crawling was employed to collect links of all the available public content of a target website through iterative access in every Web page and its hyperlinks. Scraping was used to extract specific contents of the crawled Web pages, excluding parts that are not meaningful for the system. In total, 173.02 MBs of textual data and 1,317 images of particular architectural interest were acquired.

**Table 1** List of crawled and scraped thematic websites

| Website | Description |
|---|---|
| https://www.llobregatdigital.cat/ | A news portal of L'Hospitalet with opinions and interviews with regard to culture, economy, society, health and politics |
| http://localmundial.blogspot.com/ | A personal blogspot from Manuel Domínguez that contains topics about history, society, politics and economy for the region of L'Hospitalet |
| https://elfar.cat/ | A digital archive of a monthly newspaper distributes in the broader area of L'Hospitalet whose topics cover a wide range of interests for the local residents |
| http://www.estrellalh.com/ | Proximity means of communication centered in the city of L'Hospitalet de Llobregat. An initiative of the Foment de la Informació Crítica (FIC-LH) organization, with the objective of promoting knowledge and exchange of information among L'Hospitalet residents |
| https://www.dezeen.com/ | Website of Dezeen, the world's most popular and influential architecture and design magazine, and the winner of several awards for journalism and publishing |
| https://www.archdaily.com/ | A platform to gather and share the most important information for architects who aspire to make the world a better place |
| https://www.danieldavis.com/ | Contains personal opinions of a researcher with interests in two main topics: the way technology influences architecture, and the way architecture influences people |

## Strengthening Citizens' Security

As presented in the section "Related Work", the effective analysis of the content produced in online sources (such as social media platforms) can lead to the early detection of delinquent, even criminal, behaviours taking place either in the online or in the offline world, thus negatively effecting the daily life of a society's citizens. Based on the latest advances of deep neural networks as well as on well-established statistical models, a variety of tools have been developed that enable near-real-time incident (crisis events) detection, early detection of criminal activities, as well as identification of a multiple accounts possibly managed by the same delinquent natural person.

### Crisis Event Detection

Social media constitute an important part of modern life as they encourage interaction, news dissemination, and the exchange of ideas and opinions. For instance, utilising data coming from Twitter (a very popular social media platform) can help analyse trends and events, as well as provide updates and detection of key events in close to real-time, thus even potentially lessening the impact of a crisis event.

With Deep Learning based models being the state-of-the-art for identifying crisis events in online textual resources, Convolutional, Recurrent, and, more recently, Graph Convolutional neural networks and Attention mechanisms demonstrate satisfactory performance on well-organised document collections. However, there is limited research in dealing with short, noisy texts [28, 55], present in social media.

Therefore, we developed a new technique that allows for an effective detection of crisis-related events in brief, noisy texts [56]. Our method is based on cutting-edge self-attention encoders [57] that denoise text before passing it to further processing. Overall, three self-attention-based models were created and evaluated, i.e., AD-MCNN, AD-PGRU, and Stacked-SAE, as described below.

The foundational component of our models is a Self-Attention Encoder [57], which embodies 2 sub-layers. The first is a multi-head self attention and the second is a *position-wise feed-forward* layer, i.e., a fully connected layer with shared parameters over the sequence, applied in each position. Each layer output is added with a residual connection [58] from the previous layer output followed by Layer Normalization [59] aiming at a more stable and better regularized network.

The primary idea underlying self-attention mechanisms is to let the model select which elements of a sequence are most significant, i.e., which should receive greater attention and increase their significance in the latent space. Therefore, using it as a feature extractor on the noisy text enables the discovery of important interconnections between words, while also reducing the impact of the noisy parts.

The first architecture uses the Self Attention Encoder described above as a Denoising feature extractor and forwards the output to a Multi Channel CNN [55], i.e. three parallel CNN layers operating under different kernel sizes to capture different $N$-gram combinations from the text, and a max-over-time pooling operation [60].

In this architecture, we replace the MCNN with a Parallel GRU, which includes 3 GRU units, that compress the

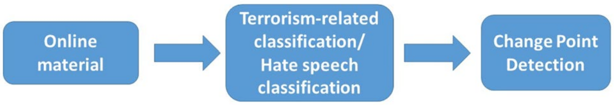

Online material
Terrorism-related classification/ Hate speech classification
Change Point Detection

**Fig. 4** Change point detection framework

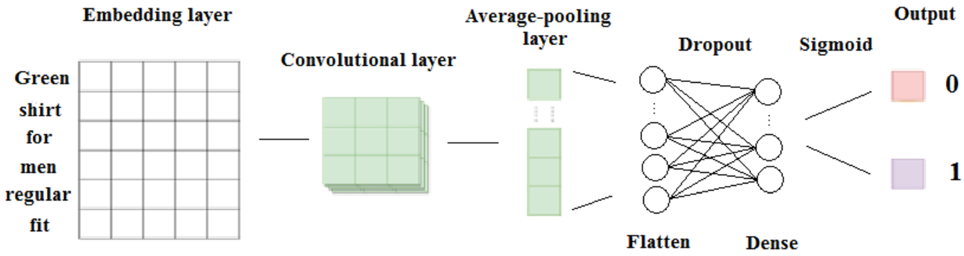


Fig. 5 Overview of the CNN-based classification model

sequence to a single vector and concatenate it before forwarding it to the classification layer. Therefore, each unit must learn a different way to interpret the denoised input sequence.

The final design comprises a stack of 4 self-attention encoders. The output layer is a fully linked layer using softmax as the activation function. We expect that this deep architecture should generate better representations and thus capture more complicated patterns.

## Change Point Detection in Terrorism-Related Online Content

In general, change point analysis in time series data enables the estimation of statistically significant changes in their structural behaviour at certain time locations. Considering the exploitation of online content from social media sources towards the extraction of valuable information for the detection and prevention of potentially suspicious content, e.g., terrorism and crime activity, the adoption of a change point detection (CPD) framework serves as a tool to analyse the temporal evolution of several indicators aiming to identify breaks that may indicate the occurrence of events where attention should be paid to.

CPD approaches are classified into two categories: online methods aiming to detect changes in real-time [61], and offline methods [62] that retrospectively identify changes taking into account past data. In our work presented in [63], we focus on offline methods and the proposed framework constitutes of two main steps. At first, the online content is classified as related to terrorism or as belonging to the hate speech category in the sense of containing expressions of aggressive behaviour. Then, appropriate time series are constructed based on the extracted indicators and change point analysis is applied to them so as to detect changes that may signal the occurrence of events of interest. The analysis takes into account multivariate time series data enabling the exploitation of possible correlations that may exist between the time series formulated by the indicators of interest. Figure 4 illustrates the overall CPD framework.

The classification part of the proposed framework has been designed to organise the obtained online content into two categories: (i) related to terrorism or not, and (ii) containing hate speech or not. The development of the two classification models is based on Convolutional Neural Networks (CNNs) using the architecture depicted in Fig. 5, inspired by [55]; for more details please check [63].

To build the terrorism-related classification model where texts are characterised as related to terrorism or not, the following datasets were used (non-English posts were disregarded):

- How ISIS uses Twitter,[6] which consists of ≈17k tweets from 100+ pro-ISIS fanboys from all over the world since the November 2015 Paris Attacks; it is considered as terrorism-related.
- Hate speech offensive tweets [64], with ≈24k labelled tweets organised into three classes: hate speech, offensive, and neither; it is considered as non-terrorism related since it was constructed by randomly retrieving content from Twitter, based on a set of hate speech related words and thus it is less likely to contain any terrorism-related content.

Overall, the ground truth dataset is split into train, test and validation sets of 37,973, 3797, and 421 samples, respectively. The overall Accuracy and F1-score are equal to 93%, while AUC equals to 99%. For the terrorism class the model achieves F1-score equal to 91%, while for the non-terrorism class the F1-score is 94%.

Focusing on the hate speech classification model, the following two datasets were used:

- A hate speech dataset that contains texts extracted from the Stormfront [65]. Overall, it consists of 1190 hate and 9462 non-hate instances.
- The above-mentioned hate speech offensive tweets dataset, which consists of ≈24k samples split in three classes: hate, offensive, and neither. The "hate" and "offensive" instances are considered as part of the hate class, while the ones labelled as "neither" as part of the non-hate class.

[6] https://www.kaggle.com/fifthtribe/how-isis-uses-twitter.

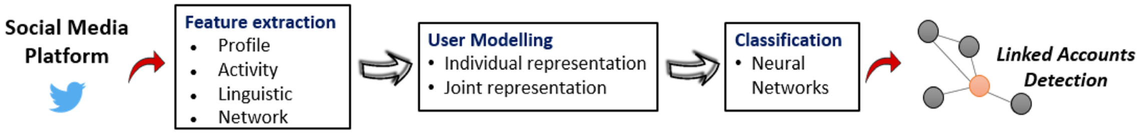


**Fig. 6** User Identity Linkage framework

Overall, the ground truth dataset consists of ≈35k samples and is split into 90% training (from which 10% is kept as validation set) and 10% test sets, maintaining the proportion of classes. Similar to the terrorism-related classification model, the overall Accuracy and F1-score equal to 93%. Focusing on the specific classes, the F1-score equals to 94% and 91% for the hate and non-hate, respectively. Finally, the overall AUC score is equal to 98%.

The CPD part of the proposed framework adopts the method presented in detail in [66]. In general, it constitutes a non-parametric approach in the change point analysis domain using the *E-devisive* algorithm, applicable either to univariate or multivariate time series to detect distributional changes; in our case the method is applied to a two-dimensional time series that is constructed based on the outputs of the two classification models about terrorism relation and hate speech containing in online content.

## User Identity Linkage

User identity linkage aims to detect accounts likely to belong to the same natural person. Towards this direction, a framework has been designed, implemented and evaluated geared to identify the linkage between online user accounts within a social media platform [44]. In particular, the developed framework considers a wide range of user characteristics (features) for representing users' online presence, while then, based on these extracted features, deep learning-based classifiers are employed to identify accounts potentially linked to the same natural person. Figure 6 outlines the overall user identity linkage framework.

Various features have been considered to model the behaviour of *each individual user*, namely:

1. Activity (A) Features related to a user's posting behaviour, such as number of mentions, hashtags, etc.
2. Linguistic (L) Features that analyse the writing style of the author of a text. A variety of linguistic features is considered, such as mean number of characters per word, vocabulary richness, discourse markers, part-of-speech.
3. Network (N) Features extracted from the social network interactions between users and include distribution (i.e., hub, authority, Eigenvector, and PageRank centralities), connection (i.e., number of triangles a node belongs to), and segmentation (i.e., clustering coefficient) metrics.

The extensive list of features considered can be found in [44].

The three above-mentioned feature categories (or sets), i.e., $S = \{A, L, N\}$, and the included features are used to model the behaviour of each individual user account. In particular, we define the feature vector for each user $u_i$ and feature category $S$ as $V_{S_{u_i}} = < f_{S_{i_1}}, f_{S_{i_2}}, \ldots, f_{S_{i_n}} >$, where $f_{S_{i_j}}$ is the $j$th feature of category $S$ for user $u_i$, and $n$ equals to the total number of included features for this category. A feature vector $V_{All_{u_i}}$ can also be created by considering all features from all three categories.

The next step involves the joint representation of user accounts to determine their potential relationship. To this end, we jointly represent the behaviour of each pair of users $u_i$ and $u_j$, $\forall i, j$, where $i \neq j$, as a feature vector of the absolute differences between the individual feature vectors of $u_i$ and $u_j$.

For the linguistic features, we also measure the similarity of two users' posts in terms of their (i) edit distance, i.e., number of changes needed to convert a text to another (Levenshtein distance is used), and (ii) semantic similarity. Semantic similarity is estimated based on a vector space model approach, whereby each word in a post is represented as a word embeddings vector; we use Word2Vec [67] to output a vector representation for each word encountered in the input texts. Given the vector representations of all words in a post, the overall vector representation of the post is derived by averaging the vectors of all its words. Finally, the set of all posts by a user, referred to as document $d$, is represented as a vector which contains the semantic center of all posts' vectors, $p$: $Sem_{center}(d) = \sum_{p \in d} vec(p)/|d|$, where $|d|$ is the number of the user's posts.

Regarding classification, a deep learning-based model is built, which combines raw text with metadata (i.e., activity, linguistic, and network features); traditional machine learning classifiers were also examined (for the corresponding results please check [44]). Combining raw text with additional behavioural evidence allows for different aspects of users' behaviour to be captured and therefore potentially leading to more effective identification of accounts that may belong to the same user. In particular, we construct a single

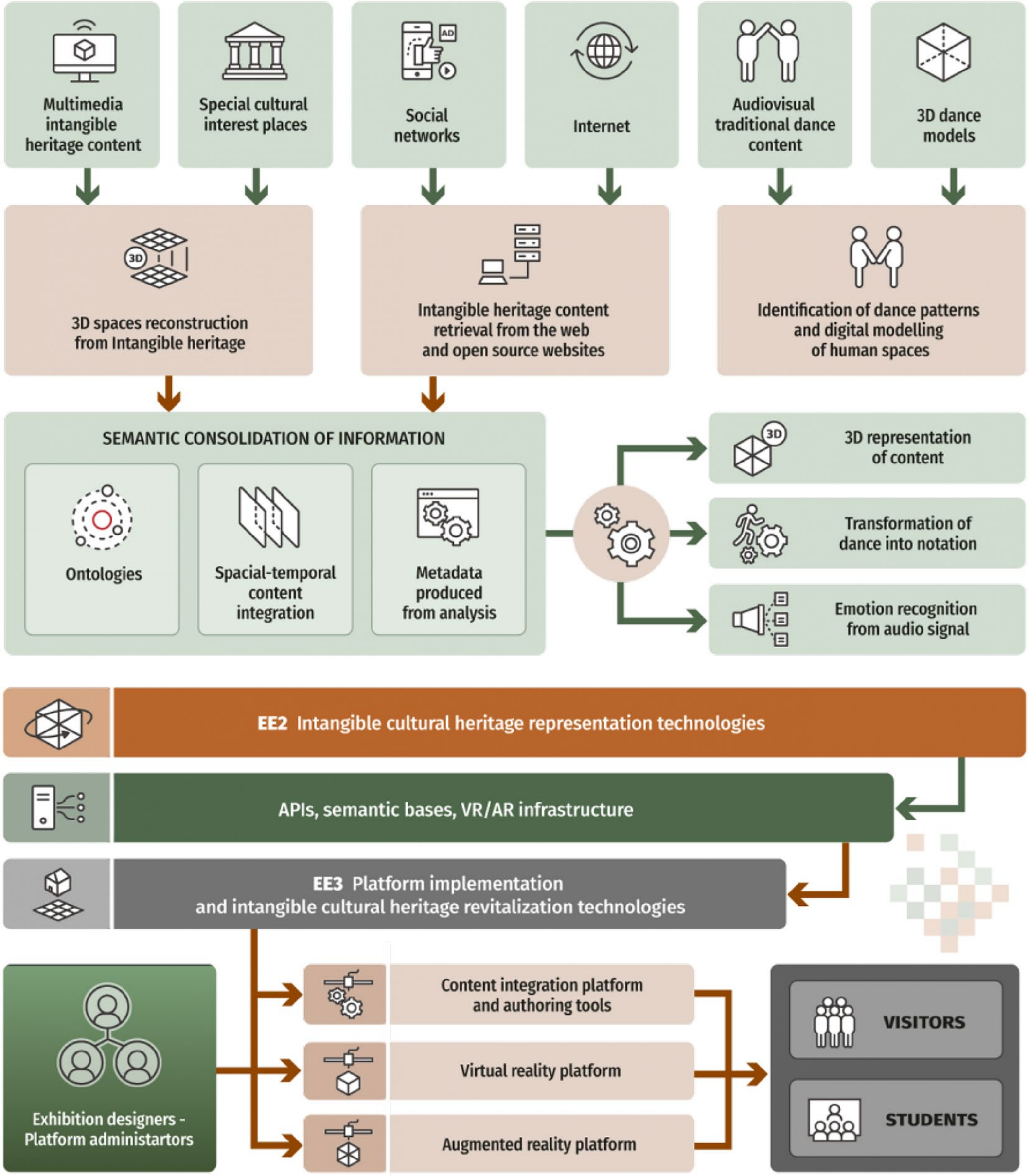

Multimedia intangible heritage content
Special cultural interest places
Social networks
Internet
Audiovisual traditional dance content
3D dance models
3D spaces reconstruction from Intangible heritage
Intangible heritage content retrieval from the web and open source websites
Identification of dance patterns and digital modelling of human spaces
SEMANTIC CONSOLIDATION OF INFORMATION
Ontologies
Spacial-temporal content integration
Metadata produced from analysis
3D representation of content
Transformation of dance into notation
Emotion recognition from audio signal
EE2 Intangible cultural heritage representation technologies
APIs, semantic bases, VR/AR infrastructure
EE3 Platform implementation and intangible cultural heritage revitalization technologies
Exhibition designers - Platform administartors
Content integration platform and authoring tools
Virtual reality platform
Augmented reality platform
VISITORS
STUDENTS

**Fig. 7** Overview of the CHROMATA system

network architecture that combines both text classification using RNNs and metadata networks before their inputs are translated into classification probabilities [44].

## Cultural Applications

In the upcoming sections ("Virtual Environments for Safeguarding Intangible Cultural Heritage" and "Preserving Cultural Heritage of Refugees") two applications are introduced with the utilization of social media data to assist in preservation of the cultural heritage.

### Virtual Environments for Safeguarding Intangible Cultural Heritage

Cultural Heritage is made up of both tangible (e.g., monuments and artifacts) and intangible (e.g., songs, performing arts, rituals, oral traditions, social practices) aspects that have been passed down from generation to generation. Preserving intangible cultural heritage (ICH) is a challenge in modern societies. To address this, the project CHROMATA [7] created a virtual reality platform [68] to gather elements of

[7] https://chromata-project.gr.

**Table 2** Thracian Folklore Dance Dataset

| Dance | No. of Videos | Duration | GBs in Storage |
|---|---|---|---|
| Gikna | 13 | 28 min of footage | 0.57 |
| Mpaintouska | 15 | 28 min of footage | 0.85 |
| Karsilamas | 16 | 40 min of footage | 1.89 |
| Hasapikos | 15 | 37 min of footage | 1.40 |

ICH as well as to create immersive virtual and augmented reality environments, capable of revitalizing and representing elements of intangible heritage with the potential of new applications on them. The high level concept of the system is illustrated in Fig. 7.

To collect the needed data of cultural interest, modules have been created that retrieve public textual and visual content from open Web sources. A module for collecting videos from YouTube have been developed around the YouTube API v3,[8] returning videos based on queries and displaying them on the Web platform with the user able to filter, inspect and download the results. The raw video files are saved on an Amazon S3 server,[9] while the metadata are saved on an internal SQL server. The total videos that were retrieved so far from social media platforms count to 513, with the video dataset of Thracian folklore dance depicted in Table 2. The main topics of the collections to gather ICH for the project are: folklore customs, actuators of customs, people participating in folklore customs, places of interest, songs and odes, and dances.

The retrieved data along with data provided by cultural institutions is used for training algorithms to extract useful information regarding the human activities conducted in a cultural framework. More precisely, the content is used for 3D Pose Estimation, Dance Recognition, Sentiment Analysis and Laban Generation.

### Preserving Cultural Heritage of Refugees

Humanitarian crises result in large-scale population movements and forced displacements, with migrants struggling to be included into the communities that welcome them. The aim of the SO-CLOSE project[10] is to promote social cohesion and combat refugee marginalization or exclusion by facilitating the encounters between people with similar life stories, via the use of innovative digital and artistic tools, with history and personal stories as central narrative threads.

For this task tools were developed to enable the collection and sharing of cultural heritage and co-creation of new cultural materials with and for refugees. Those tools are available to the curators and the public from a multimodal Web platform called Memory Center Platform.[11] To create the visual and textual datasets needed, different modules were created, accessible through API endpoints, collecting open access content from Twitter, YouTube and selected websites.

The Twitter module is developed around Twitter API v1.1,[12] creating collections in the database for textual analysis tasks. This service retrieves real-time public Twitter posts based on keywords, key-phrases and mentions of a user-defined collection. For the purposes of the project, Twitter module crawling iteration lasted one month starting from May, 2022, for predefined thematic collections of keywords resulting in vast amounts of data. Sample collections used in the project are depicted in Table 3.

Similar to the previous module, the YouTube module collects video entities or metadata, based on thematic queries to create collections in the database. It is developed around YouTube v3 API, retrieving and storing video metadata according to specified keywords and number of videos the user wishes to review. Only videos with Creative Commons licenses are displayed and collected in this manner, with the option to download the video file locally or in the project database.

Finally, a focused crawler and scraper was developed to collect data from thematic websites. The websites selected are shown in Table 4 and the decision to crawl them was based on (i) legal issues on crawling/content reusability, (ii) abundance of multimedia content, (iii) well structured templates/compatibility. The module crawls the website iterating each Web page and its hyperlinks collecting relevant data and storing them in the database.

The collected data of all the above tools are stored in a MongoDB database and are displayed on the Memory Center Web platform where users can view, sort and filter out the results. An example interface of the platform is displayed in Fig. 8.

## Disaster Management

In these times disasters that occur all over the world are being mentioned swiftly in the social media platforms. Social media play an important role in information gathering and has been proven to be a valuable source of information for early warning tools during a disaster.

[8] https://developers.google.com/youtube/v3/docs/search/list.

[9] https://aws.amazon.com/s3/.

[10] https://so-close.eu.

[11] https://mcpwebstart.net/.

[12] https://developer.twitter.com/en/docs/twitter-api/v1.

**Table 3** SO-CLOSE collections details

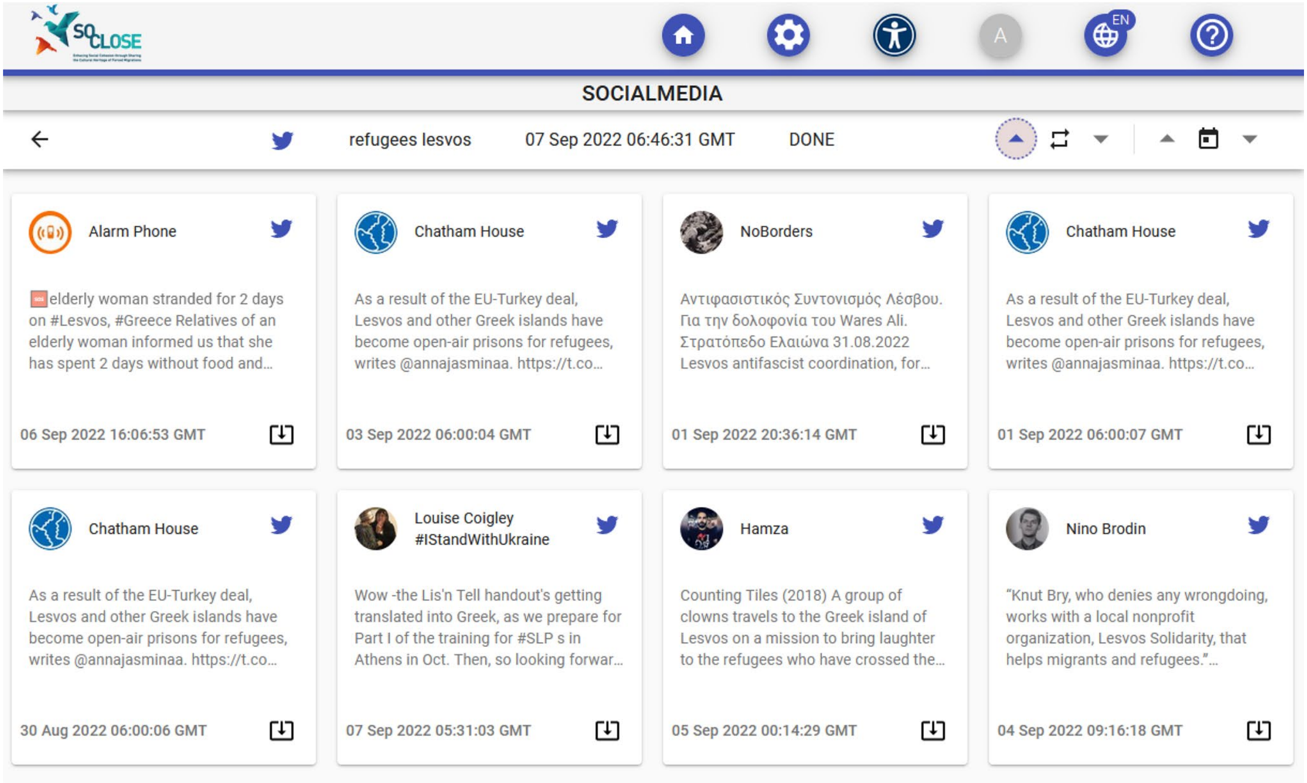
| Collection Title | Keywords/Key-phrases included | Tweets | YouTube Metadata | Videos |
|---|---|---|---|---|
| Definitions of actors in the process of social cohesion | Refugee, immigrant, foreigners, asylum seekers, forced displacement | 116,174 | 30 | 8 |
| Needs and rights of refugees | Refugee needs, refugee rights, right to citizenship, refugee documents, right to work, right to food, access to health, access to education, access to housing | 352,012 | 36 | 10 |
| Discrimination in host society | Social stigma, social exclusion, racism, racial stereotypes, discrimination, steal jobs, blaming refugees | 15,282 | 47 | 10 |
| Gender | Transphobia refugee, LGBTQIA+ refugee, refugee women rights, transgender refugee, homophobia refugee, gay refugee | 496 | 0 | 0 |
| Cultural heritage | Cultural backgrounds, refugee shared heritage, refugee memorial heritage, workshop refugee, project refugee, refugee inclusion, refugee culture, refugee tradition, documentary refugee, poem refugee, refugee cinema, refugee storytelling, refugee recipes, refugee science, refugee language | 3,233 | 83 | 15 |
| Emotions | Empathy refugee, compassion refugee, anger refugee, refugee nostalgia, fear refugee, emotional heritage | 1,170 | 47 | 6 |
| Geography | Refugee Poland, refugee Italy, refugee Greece, refugee Spain, refugee Catalunya, crossing borders | 5,233 | 46 | 1 |
| Trauma | Refugee support group, refugee violence, refugee trauma, vulnerable refugee | 1,980 | 40 | 10 |
| Society | Social integration, social inclusion, racist society, refugee integration, refugee community, foreign community, social cohesion, refugee acceptance, refugee empowerment, refugee working class, intercultural refugee, refugee oppression, refugee resistance | 28,143 | 95 | 20 |
| Religion | Muslim minority, Islam refugee, hijab refugee, refugee mosque, Ramadan refugee | 7,156 | 48 | 3 |
| Tools | Refugee phones, refugee camera, refugee virtual, refugee twitter, refugee Facebook, refugee Instagram, refuge social media | 59,101 | 50 | 10 |
| History and memory | European memory, memorial heritage, civil war, colonialism, decolonise, collective memory, dictatorship, exile | 565,385 | 50 | 7 |
| Approach in working with refugees | Refugee victimhood, refugee assimilationism, ethnicization, folkorization, exoticism, eurocentric | 5,246 | 48 | 10 |
| Displacement | Refugee homeland, refugee fatalism, refugee journey, country nostalgia | 2,238 | 50 | 6 |

**Table 4** List of crawled and scraped thematic websites

| Website | Short description | Focus |
|---|---|---|
| https://www.amnesty.org/en/ | World's largest human rights movement | News Reporting/Policy |
| https://www.digitalmeetsculture.net | Portal for gathering information about world digital culture | News Reporting on Digital Culture |
| http://cultural-opposition.eu | Project funded by EU | Cultural opposition/Socialism in Eastern Europe |

In order to utilize the vast information available on social media, we implemented an application that retrieves data from the social media platform of Twitter and aims to resolve the challenges that appear on it. The application retrieves tweets in real time about fires, earthquakes and terror attacks, and timely detects events for early warning tools that can assist first responders in decision making.

The overall architecture of our application's framework is illustrated in Fig. 9. The core component of the framework is a Twitter Crawler, which establishes a connection to the Twitter Streaming API and collects continuously new posts (tweets) directly from Twitter, based on some user-defined search criteria (e.g., keywords and Twitter accounts). The collected tweets are analyzed using: (i) the Verification API, to estimate the reliability of the tweet and the probability that it contains misinformation, (ii) the Relevance Estimation API (see the section "Relevance Estimation"), to filter out irrelevant/noisy tweets, and (iii) the Localization



SOCIALMEDIA

refugees lesvos 07 Sep 2022 06:46:31 GMT DONE

Alarm Phone

🟥elderly woman stranded for 2 days on #Lesvos, #Greece Relatives of an elderly woman informed us that she has spent 2 days without food and…

06 Sep 2022 16:06:53 GMT

Chatham House

As a result of the EU-Turkey deal, Lesvos and other Greek islands have become open-air prisons for refugees, writes @annajasminaa. https://t.co…

03 Sep 2022 06:00:04 GMT

NoBorders

Αντιφασιστικός Συντονισμός Λέσβου. Για την δολοφονία του Wares Ali. Στρατόπεδο Ελαιώνα 31.08.2022 Lesvos antifascist coordination, for…

01 Sep 2022 20:36:14 GMT

Chatham House

As a result of the EU-Turkey deal, Lesvos and other Greek islands have become open-air prisons for refugees, writes @annajasminaa. https://t.co…

01 Sep 2022 06:00:07 GMT

Chatham House

As a result of the EU-Turkey deal, Lesvos and other Greek islands have become open-air prisons for refugees, writes @annajasminaa. https://t.co…

30 Aug 2022 06:00:06 GMT

Louise Coigley #IStandWithUkraine

Wow -the Lis'n Tell handout's getting translated into Greek, as we prepare for Part I of the training for #SLP s in Athens in Oct. Then, so looking forwar…

07 Sep 2022 05:31:03 GMT

Hamza

Counting Tiles (2018) A group of clowns travels to the Greek island of Lesvos on a mission to bring laughter to the refugees who have crossed the…

05 Sep 2022 00:14:29 GMT

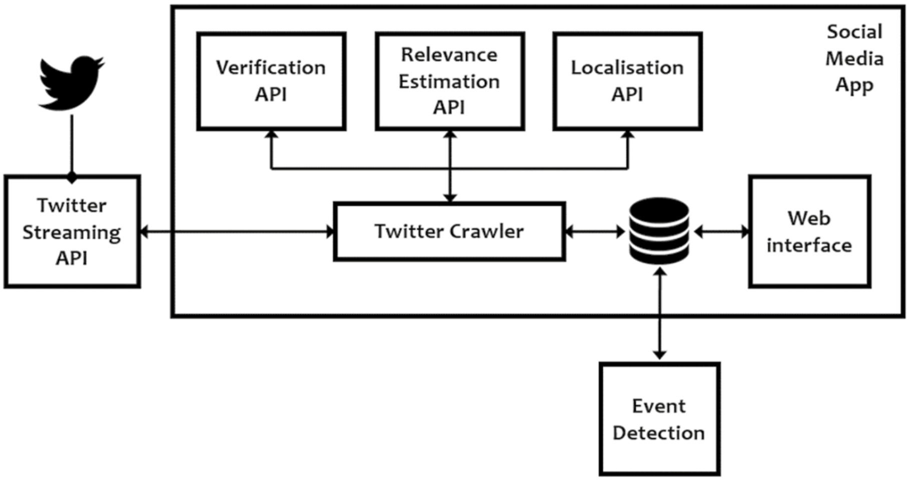
Nino Brodin

"Knut Bry, who denies any wrongdoing, works with a local nonprofit organization, Lesvos Solidarity, that helps migrants and refugees."…

04 Sep 2022 09:16:18 GMT

**Fig. 8** Display example of collected tweets in Memory Center Platform

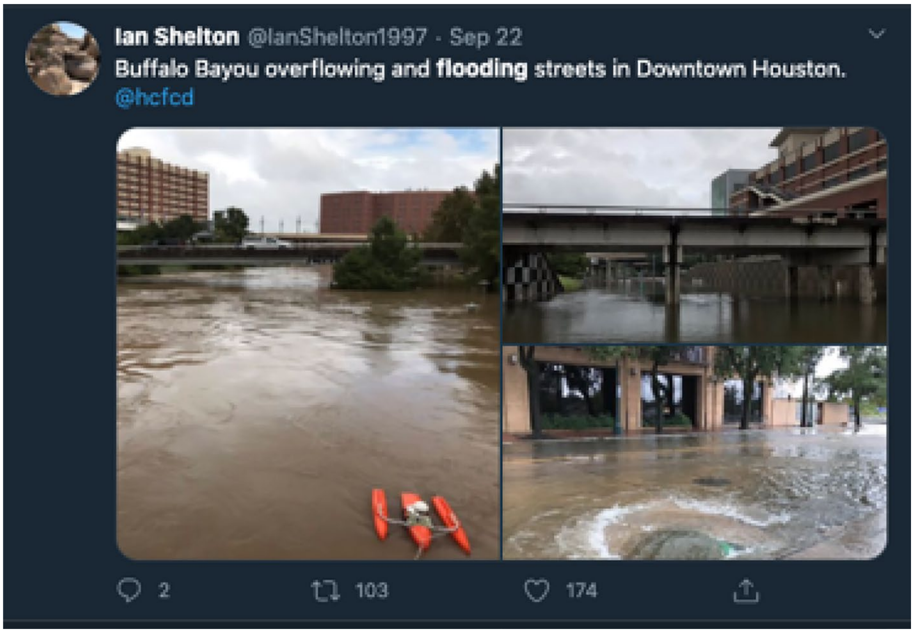
**Fig. 9** Architecture of the Social Media App framework



API, to detect and extract the locations from the tweet's text and associate them to coordinates provided by the OpenStreeMap API.[13] The results of each analysis and the tweet's metadata are stored in a MongoDB database. Then, the collected and analyzed tweets are displayed on a Web interface[14] where the end users are able to view and filter the posts. The interface can also be used to collect manual annotation by users w.r.t. their relevance to the involved disasters, so as to train the relevance estimation model. Finally, the Event Detection component (see the section "Event

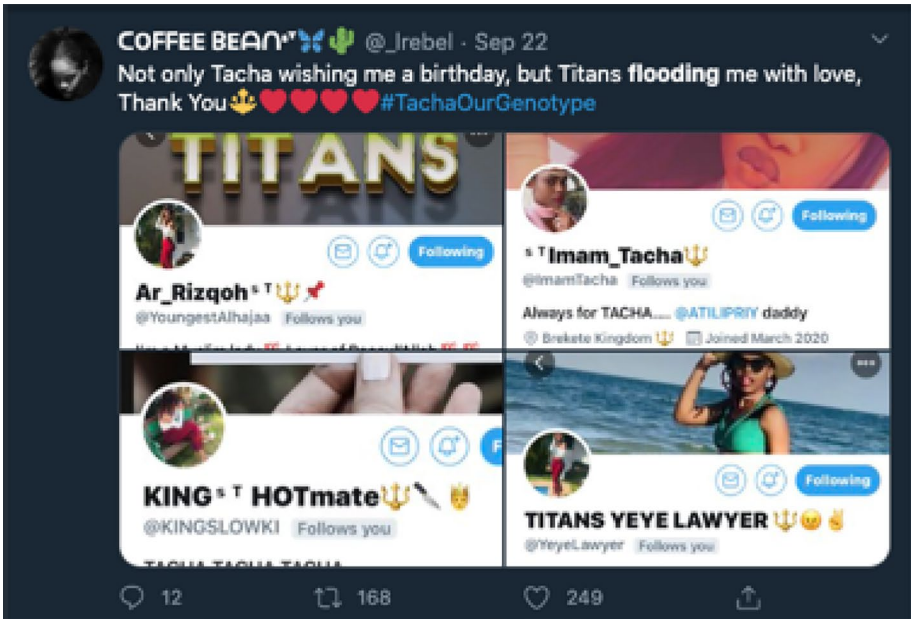
[13] https://wiki.openstreetmap.org/wiki/API.

[14] https://socialmedia-server-m4d.iti.gr/ingenious/relevancy-annotation.html.



Ian Shelton @IanShelton1997 · Sep 22
Buffalo Bayou overflowing and flooding streets in Downtown Houston. @hcfcd

**Fig. 10** Example of relevant tweet



COFFEE BEAN @_Jrebel · Sep 22
Not only Tacha wishing me a birthday, but Titans flooding me with love, Thank You 🔱❤️❤️❤️❤️ #TachaOurGenotype

**Fig. 11** Example of irrelevant tweet

Detection") examines the analyzed tweets and produces warnings for potential incidents.

### Relevance Estimation

Distinguishing relevant information from a large amount of social media data in the middle of a disaster event is a task that can assist the management of the valuable manpower to be allocated more efficiently on crucial missions. To accommodate this task, a relevance estimation service based on a machine learning approach is proposed to identify whether a post retrieved from social media belongs to a specific scenario.

In our scenario (see the section "Disaster Management"), relevance estimation aims to separate relevant posts (Fig. 10) from irrelevant (Fig. 11) and thus filter out unrelated tweets because of the metaphorical use of incident-related keywords. The first step of the training of the proposed machine learning classifier is the data preprocessing by removing URLs, emojis, mentions ("@"), punctuation and all non-characters from the text of the tweets. Secondly, for the process of the feature extraction

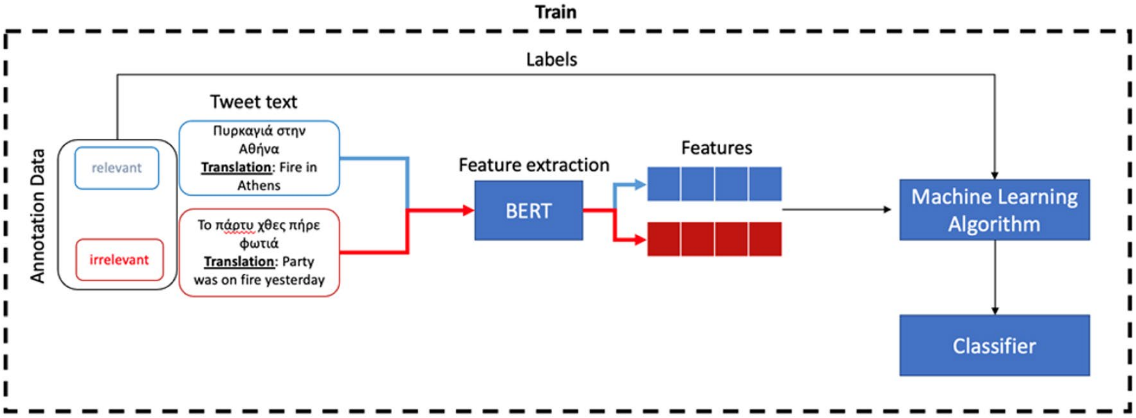


Fig. 12 Architecture of the relevance estimation methodology

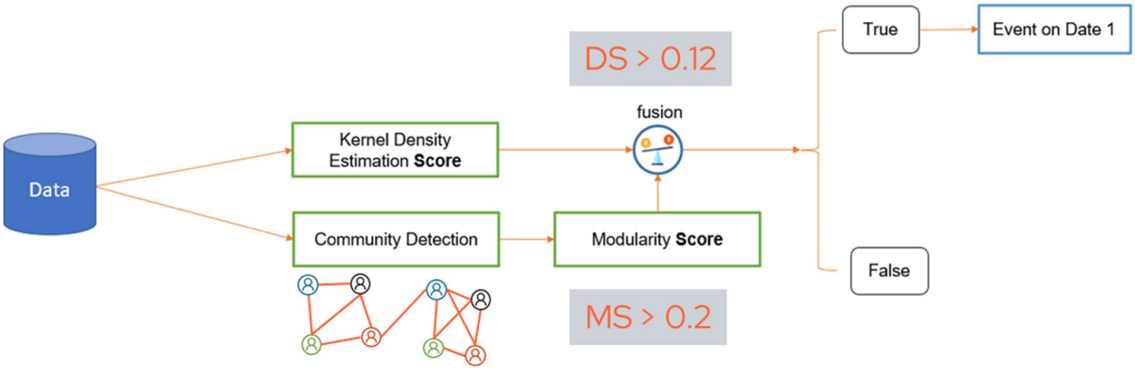


Fig. 13 Architecture of the event detection methodology [72]

we implement a state-of-the-art Bert [69] that enables our model to grasp the context of a word based on its surroundings. After the feature extraction, tweets texts are converted into a numerical representation in the form of a vector in a format suitable to feed a machine learning algorithm.

Moreover, for training our model we use the pairs of the feature sets extracted from texts and the labels of each tweet as "1" for relevant and "0" for irrelevant (created with manual annotation), in order to produce a classification model. Logistic Regression is considered for the classification and predicts the probability of an instance belonging to the default class, i.e., 1/relevant. The complete architecture of the relevance estimation methodology can been seen in Fig. 12.

Finally, once the model is trained, the same feature extractor can be applied to transform unseen text into a vector, which can be then fed into the classification model to get predictions (i.e., irrelevant or relevant). More information about the evaluation of the proposed relevance estimation model can be found in the section. "Estimating Relevance to Fire Disaster".

## Event Detection

In this section, we describe the methodology in which the huge stream of single posts collected by the Twitter Crawler (see the section "Disaster Management") are grouped into a more compact form of information that is useful for the end user. This can be achieved using event detection methods, in order to find events through the data and provide higher knowledge to the users.

The event detection methodology that is shown in Fig. 13 involves a novel three-step process that fuses two modalities. Firstly, the tweets are retrieved from the MongoDB database for a specific time period. Secondly, we calculate the Density Score (DS) by applying the Kernel Density Estimation (KDE) [70] with the utilization of the tweets timestamps of the previous step. KDE takes into account not only the number of Twitter posts for the examined time frame but also the sparsity and density of their publication time. In parallel, a Community Detection [71] algorithm runs to compute the Modularity Score (MS). Community detection is a graph-based method to discover communities of Twitter users with similar properties (e.g., discussing same topic). In the third

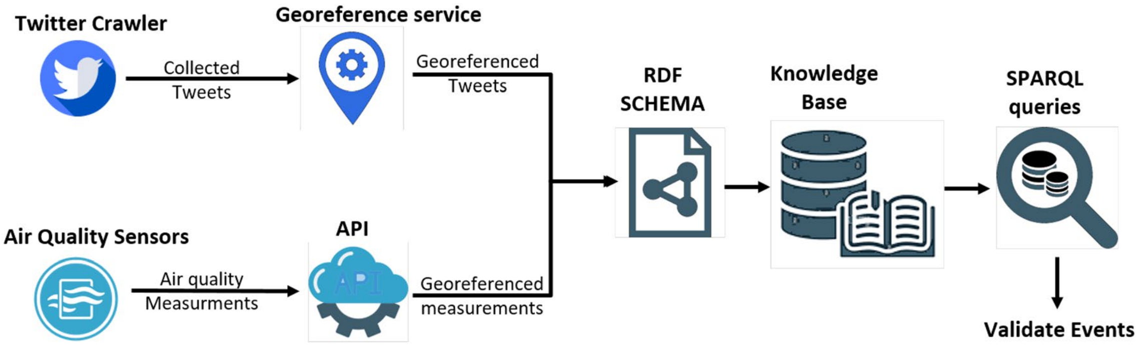


**Fig. 14** Architecture of the air quality validation framework

and last step, a fusion method considers the previous two scores and decides if an event exists in the respective time period. For additional information on the event detection framework the reader is referred to [72].

The event detected by the aforementioned methodology shares some attributes: (i) the timestamp when the potential incident was identified, (ii) the location it concerns, and (iii) the group of the tweets posted by different users that constitute the event. Further insights are also generated for a better understanding of the incident by the users (e.g., first responders), by removing the stop-words and extracting the top-ten most frequent words from the text of the tweets of the event. Information about the results of the event detection method (in particular for the disaster of fire) can be found in the section. "Detection of Fire-Related Events".

### Creeping Crisis

Event detection research in social media has been extensively focused in the early detection of disasters, which are mostly natural or human-caused disasters that occur without warning (e.g., fires, earthquakes, terror attacks). However, apart from sudden crisis events, there are also creeping crisis incidents that are a threat to life-sustaining systems and evolve over time and space, foreshadowed by precursor events [73], i.e., air quality. These crises are subtle and have less evidence than sudden crisis events, but are equally life-threatening. However, here arises the issue whether social media can be exploited in order to discover such subtle events.

In [74] we introduced a framework (illustrated in Fig. 14) for the validation of automatically georeferenced social media, in particular from Twitter in German language, for the assessment of air quality in urban areas. The initial retrieved tweets are georeferenced with a state-of-the-art Named Entity Recognition (NER) implementation and then the geolocated tweets are mapped to an RDF schema, in order to support SPARQL queries[15] for linking the social media data with real measurements from air quality sensors and perceive the correlations between them. The dataset and results of this research are discussed later in the section "Validation of Humans as Sensors for Air Quality".

### Water Management

Social media are already a game changer in natural disaster management [75, 76] and their value can be transferred to a different domain, such as water safety and security. As society evolves and social media are an integral part of it, water utilities should also evolve and consider them. Nowadays, many people choose to report water-related issues and crises (e.g., water quality issues, floods, droughts) that occur in their region via social media platforms.

Monitoring of social media can play a decisive role and assist in decision-making. Water utility operators, by taking into consideration the social media as a complementary source to their traditional means, will save time with the automatic retrieval of relevant posts by crawlers, instead of manual search. Also, crawlers offer better coverage by retrieving every single social media post that matches the search criteria, while AI-based techniques can improve the incoming information by filtering out noise and geotagging the relevant social data. Lastly, the detection of events covered by social media can assist decision-makers with the efficient allocation of human resources. In this way, human sensing from social media data will open a new channel of incoming information that will enhance the situational awareness for issues that concern the quality of water and support the creation of social awareness in a water distribution network.

It is also worth mentioning that floods and droughts can also affect the quality of water. During all phases of floods there are plenty of tweets mentioning warnings for bad

[15] https://www.w3.org/TR/rdf-sparql-query/.

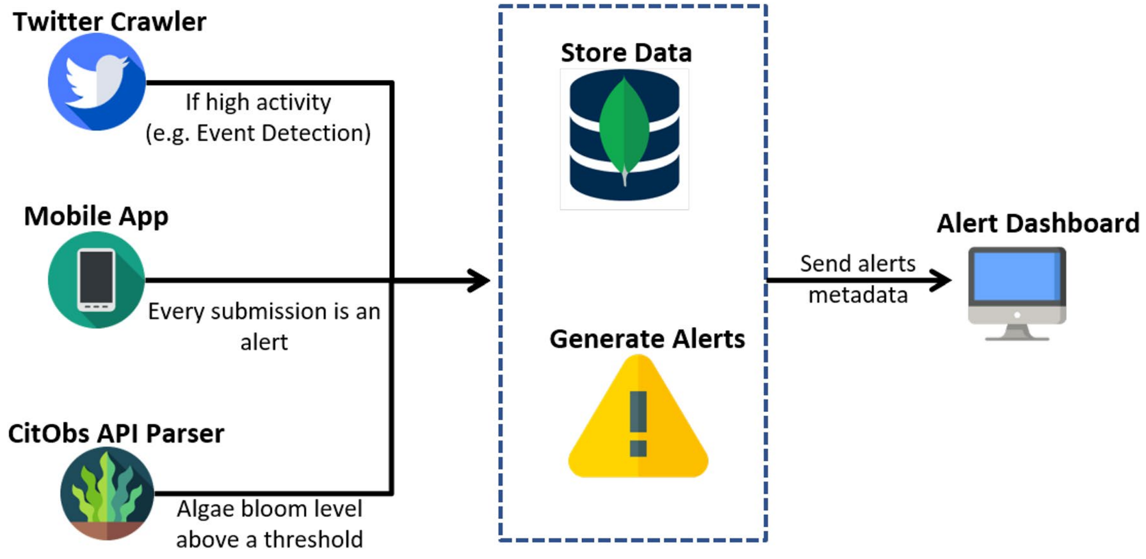


**Fig. 15** Workflow of the Alerting Module framework

weather conditions, info during the disaster and the result of the aftermath. On the other hand, tweets about droughts are harder to find, as there is no specific starting point to take as a reference for the identification of a drought event.

In the following the section "Detection of Water Quality Incidents" we describe three event detection techniques that we have utilized for recognising water quality issues from social media, and in the section "Additional Sources of Social Data" we introduce an alerting module that integrates human sensing from three different sources of citizen observation in order to create alerts for water pollution problems.

### Detection of Water Quality Incidents

In this section, we describe three different event detection techniques that can be used on social media streams in order to identify incidents that concern the quality of water.

Specifically, the three well-established event detection algorithms are Z-Score [77], STA/LTA [78] and DBSCAN [79]. These algorithms are ideal for our case, as they focus on the number of posts in relation to time, have no need for training data (it is very hard to have a Ground Truth of all occurred water-related events), and they are easy to implement.

The Z-score method is a statistical metric that, assuming a Gaussian distribution, indicates how many standard deviations away a given observation is from the mean. The STA/LTA method aims to discover the peaks (days with lots of tweets) and differentiate them from troughs (days with low amount of tweets) in social media activity. The DBSCAN method has the objective to cluster the tweets into events by grouping them based on their publication date on Twitter and their geographical location, and leave out the noise (tweets that do not refer to events).

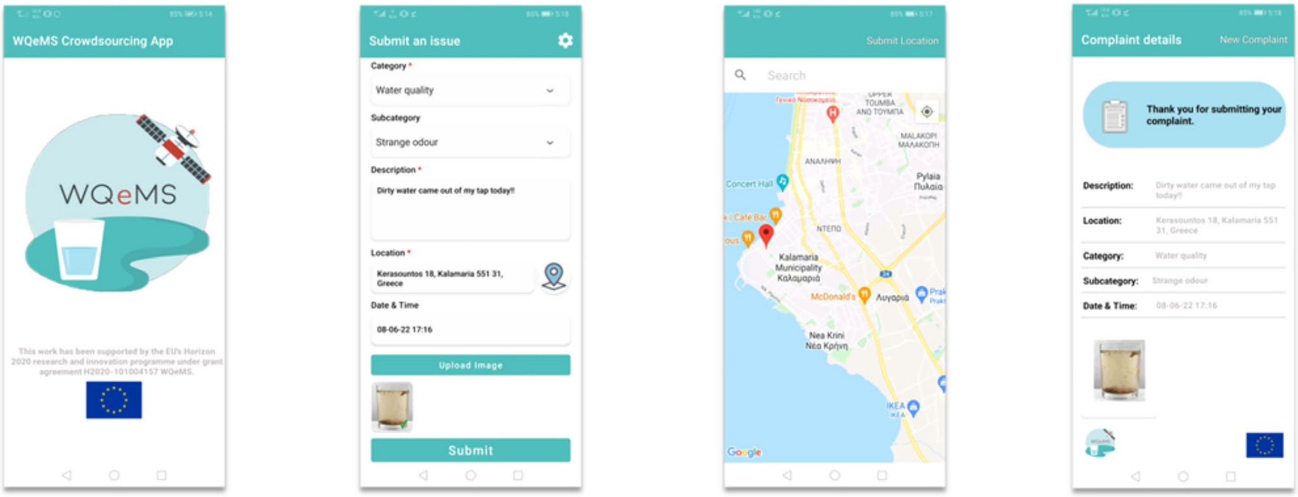
After the implementation of these methodologies and their application on a Twitter dataset, the results were possible events of water issues. The next step was to identify which of these events occurred in real life and to mark them as relevant or irrelevant based on the tweet's context. The evaluation of the three methods as well as the dataset are presented in the section. "Detection of Water-Quality-Related Events", while for more information about this work the reader is referred to [80].

### Additional Sources of Social Data

Reporting or identifying crisis events through social media platforms is often an arduous task. Since these platforms are not focused on identifying and locating hazards in order to timely alert the authorities and the first responders, there is a need for additional tools as well as additional sources of social data that will provide event detection and warning functionalities for the interested parties.

In this section, we propose the Alerting Module framework (Fig. 15) that has been developed to incorporate additional crowdsourcing information from the citizens, distinguish water-related events (e.g., debris, oil spills, algal blooms) and visualize these incidents in a dedicated dashboard. This module aims to explore the utilization of the widely used social media to obtain crowdsourced data that are relevant to water quality, but at the same time to enhance awareness with other sources of information, by combining data from three different sources.

The first source is data being collected from Twitter via a crawler that establishes an open connection to Twitter



This work has been supported by the EU's Horizon 2020 research and innovation programme under grant agreement H2020-101004157 WQeMS.

Category *: Water quality

Subcategory: Strange odour

Description *: Dirty water came out of my tap today!!

Location *: Kerasountos 18, Kalamaria 551 31, Greece

Date & Time: 08-06-22 17:16

Thank you for submitting your complaint.

Description: Dirty water came out of my tap today!!

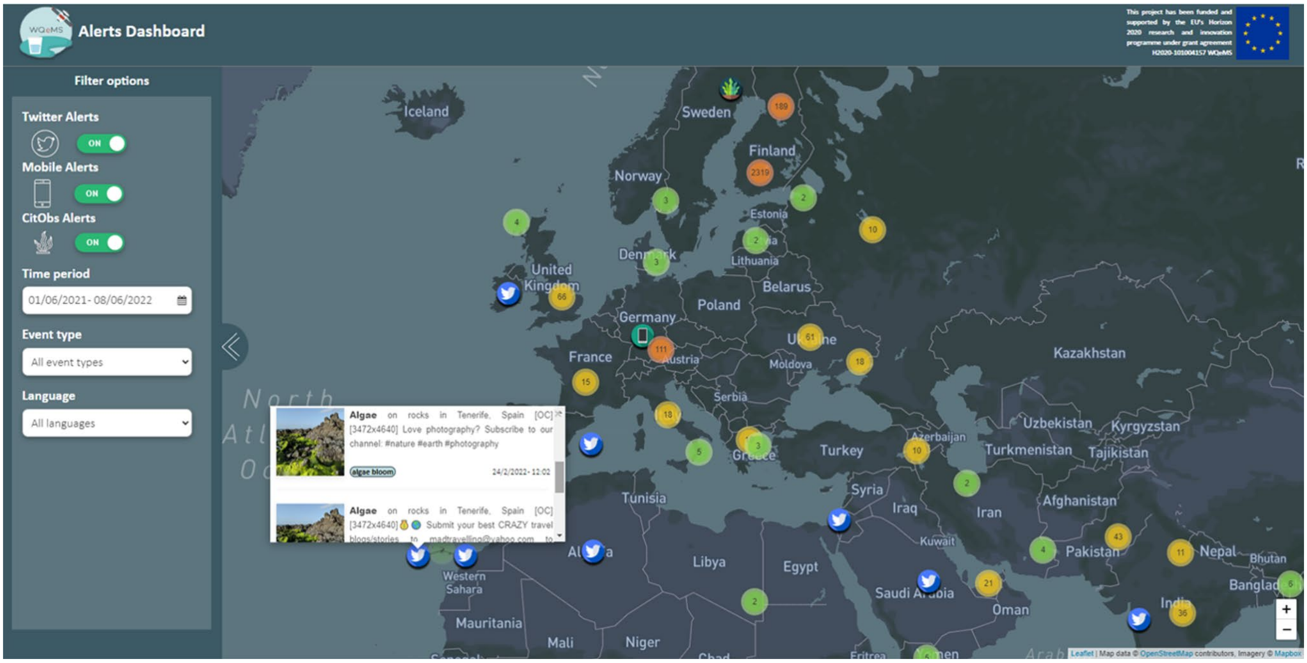
Location: Kerasountos 18, Kalamaria 551 31, Greece

Category: Water quality

Subcategory: Strange odour

Date & Time: 08-06-22 17:16

**Fig. 16** Screenshots of the Crowdsourcing Mobile App



This project has been funded and supported by the EU's Horizon 2020 research and innovation programme under grant agreement H2020-101004157 WQeMS

Twitter Alerts: ON

Mobile Alerts: ON

CitObs Alerts: ON

Time period: 01/06/2021- 08/06/2022

Event type: All event types

Language: All languages

Algae on rocks in Tenerife, Spain [OC] [3472x4640] Love photography? Subscribe to our channel: #nature #earth #photography

algae bloom 24/2/2022- 12:02

Algae on rocks in Tenerife, Spain [OC] [3472x4640] Submit your best CRAZY travel blogs/stories to madtravelling@yahoo.com to

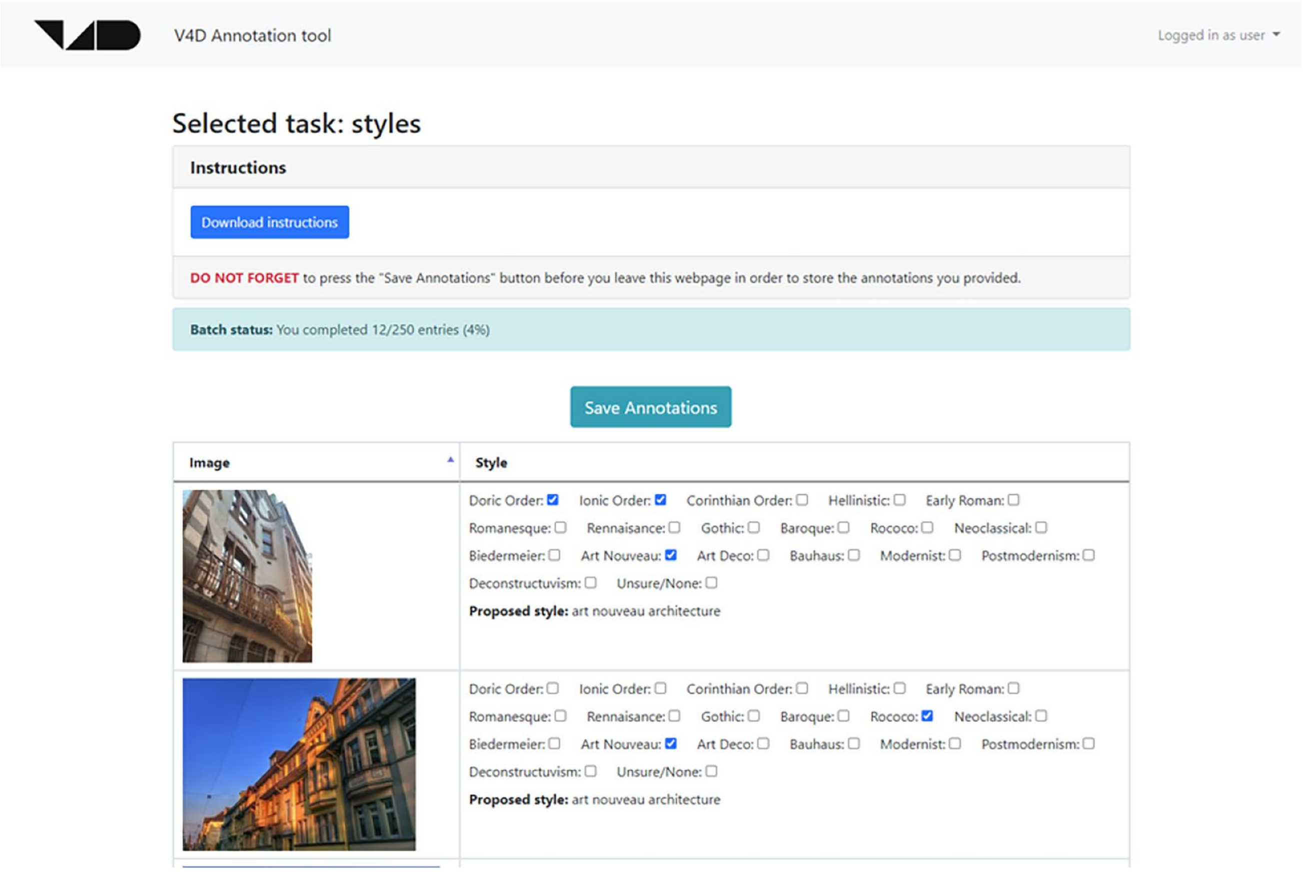
**Fig. 17** A screenshot of the Alerts Dashboard

API and retrieves in real time tweets based on user-defined keywords and selected accounts. The second source is information posted through the Crowdsourcing Mobile App, a custom Android application that we have developed in order to offer an intuitive way for reporting water-related issues. Users are able to submit a problem, by filling in a simple form with the description and the location of the water issue, while an associated photo can be attached; screenshots can be seen in Fig. 16). The third source of data is SYKE's CitObs API[16] that specifically concerns citizen

**Table 5** The architectural styles label set

| | | | |
|---|---|---|---|
| Doric Order | Ionic Order | Corinthian Order | Hellenistic |
| Early Roman | Romanesque | Gothic | Renaissance |
| Baroque | Rococo | Neoclassical | Modernist |
| Art Deco | Art nouveau | Bauhaus | Postmodernism |
| Deconstructivism | Biedermeier | | |

[16] https://www.ymparisto.fi.



V4D Annotation tool

Selected task: styles

**Instructions**

**DO NOT FORGET** to press the "Save Annotations" button before you leave this webpage in order to store the annotations you provided.

**Batch status:** You completed 12/250 entries (4%)

| Image | Style |
|---|---|
| | Doric Order: ☑ Ionic Order: ☑ Corinthian Order: □ Hellinistic: □ Early Roman: □<br>Romanesque: □ Rennaisance: □ Gothic: □ Baroque: □ Rococo: □ Neoclassical: □<br>Biedermeier: □ Art Nouveau: ☑ Art Deco: □ Bauhaus: □ Modernist: □ Postmodernism: □<br>Deconstructuvism: □ Unsure/None: □<br>**Proposed style:** art nouveau architecture |
| | Doric Order: □ Ionic Order: □ Corinthian Order: □ Hellinistic: □ Early Roman: □<br>Romanesque: □ Rennaisance: □ Gothic: □ Baroque: □ Rococo: ☑ Neoclassical: □<br>Biedermeier: □ Art Nouveau: ☑ Art Deco: □ Bauhaus: □ Modernist: □ Postmodernism: □<br>Deconstructuvism: □ Unsure/None: □<br>**Proposed style:** art nouveau architecture |

**Fig. 18** V4Design Annotation Tool

observations about algae blooms over Finland (to utilize it, we have implemented a dedicated parser).

Subsequently, the Alerting Module monitors the posts collected by the Social Media Crawler, the reports submitted by the Crowdsourcing Mobile App and the algae bloom citizen reports from CitObs API and produces alerts when, a significant number of tweets and mobile reports accumulates in a certain amount of time in specific geographic region or the algae bloom values from a specific region in Finland is above a specified threshold. The generated crowdsourced alerts from the different sources are forwarded and visualized on the Alerts Dashboard[17] (Fig. 17), a user-friendly Web interface that displays alerts on a map as pop-ups and offers filtering capabilities.

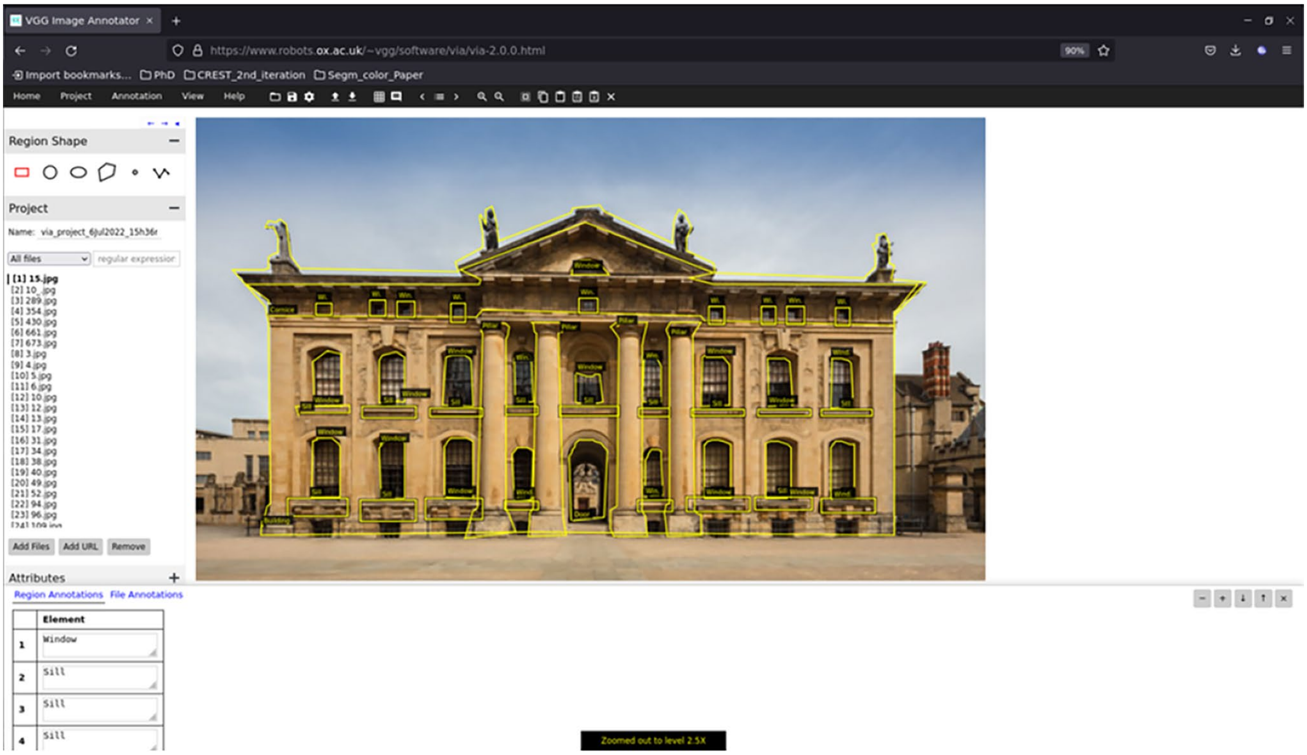
## Representative Results

In this section we illustrate some representative outcomes and some indicative results of the evaluation of the methodologies described in the section "Overview of Our Approaches in Social Sensing" in order to discuss their effectiveness.

### Dataset Creation for Architectural Style Recognition and Building Localization

The data collection module developed as part of the V4Design system (described in the section. "Reuse and Repurpose of Multimedia for Indoor and Outdoor Designs") has also been used to create datasets with ground truth annotations for visual analysis research tasks. Specifically, images from Flickr have been automatically gathered to support the training of architectural style recognition and building localization models. In the former, a classifier takes the visual data from a picture of a facade or other part of a building, and its

[17] http://m4d-apps.iti.gr:8007/WQeMS_Alerts_Dashboard.



Name: via_project_6Jul2022_15h36r

All files | regular expression

[1] 15.jpg
[2] 10_.jpg
[3] 289.jpg
[4] 354.jpg
[5] 430.jpg
[6] 661.jpg
[7] 673.jpg
[8] 3.jpg
[9] 4.jpg
[10] 5.jpg
[11] 6.jpg
[12] 10.jpg
[13] 12.jpg
[14] 13.jpg
[15] 17.jpg
[16] 31.jpg
[17] 34.jpg
[18] 38.jpg
[19] 40.jpg
[20] 49.jpg
[21] 52.jpg
[22] 94.jpg
[23] 96.jpg
[24] 109.jpg

| | Element |
|---|---|
| 1 | Window |
| 2 | Sill |
| 3 | Sill |
| 4 | Sill |

**Fig. 19** VGG Annotator

associated metadata, as input, and classifies it as one of a list of architectural "styles" or "genres" (e.g., Ionic, Baroque). In the latter, an algorithm takes the visual data from a picture of a facade of a building as input, and uses edge detection and contrast detection to create a mask separating the facade from fore- and background. A second classifier algorithm creates certainty values for the different planes on the mask, identifies different smaller planes within the mask, and classifies them as one of a list of building segments (e.g., windows, doors).

For both activities, Flickr was searched using a set of architecture styles (Table 5). The Flickr results were combined with the results from other online sources (which are not considered as social media and are out of the scope of this work) and the merged dataset was annotated with the help of Web-based annotation tools and manual input from architecture experts.

For the architectural styles recognition dataset, a custom annotation tool was developed (Fig. 18). There, images of architectural structures are presented and the annotators define the architectural style (or styles) that the structures belong to. Annotators are also provided with detailed instructions of the annotation process, as well as a proposed style for each image which corresponds to the query that brought it.

For building localization, the online software "VGG Image Annotator" [81] (Fig. 19) was selected. Each user was given a set of images that include buildings of various architectural styles presented from several viewing angles. They were asked to mark the whole building, locate all the possible façade elements within each image and label them accordingly. In the end of the process, a file was extracted that contained all the essential details about the polygons of the annotation masks and the corresponding type of façade they represent.

The architectural styles dataset contains in total 9,037 annotations, whereas the building localisation one 250 annotations. More information as well as details on how to access the datasets are available online[18].

## Detecting Crime-Related Crisis Events

In order to evaluate the proposed architectures (see the section "Crisis Event Detection") we used CrisisLexT26 [23], since, as already mentioned, it is publicly available and widely used in related work. It contains 26 different crisis events, such as floods, earthquakes, bombings, etc., that took place in 2012 and 2013. Overall, it consists of ≈ 28k tweets, with ≈1,000 posts per crisis event. The labels for each tweet concern its: (i) *Informativeness*, whether it is related to a specific crisis or not, (ii) *Information Source*,

[18] https://v4design.eu/v4design-dataset-2/.

**Table 6** Experimental results

| | Imbalanced | | | | Balanced | | | |
|---|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1-score | AUC | Precision | Recall | F1-score | AUC |
| Binary Classification | | | | | | | | |
| MCNN | **0.841** | 0.772 | 0.800 | 0.921 | 0.798 | 0.793 | 0.793 | 0.918 |
| MCNN-MA | 0.772 | 0.771 | 0.770 | 0.888 | 0.691 | 0.768 | 0.711 | 0.871 |
| Stacked-SAE | 0.821 | 0.799[a] | 0.808[a] | 0.915 | **0.809** | 0.784 | 0.793 | 0.910 |
| AD-PGRU | 0.835 | 0.799[a] | 0.814[a] | **0.927**[a] | 0.808 | 0.802 | **0.803**[a] | 0.921 |
| AD-MCNN | 0.834 | **0.802**[a] | **0.816**[a] | 0.925[a] | 0.804 | **0.805** | 0.802[a] | **0.923**[a] |
| Multi-class Classification | | | | | | | | |
| MCNN | **0.671** | 0.627 | 0.640 | 0.913 | 0.624 | **0.648** | 0.632 | **0.910** |
| MCNN-MA | 0.616 | 0.589 | 0.598 | 0.883 | 0.561 | 0.577 | 0.563 | 0.873 |
| Stacked-SAE | 0.644 | 0.630[a] | 0.633 | 0.906 | 0.622 | 0.636 | 0.626 | 0.900 |
| AD-PGRU | 0.648 | 0.637[a] | 0.638 | 0.910 | **0.627** | 0.640 | 0.630 | 0.909 |
| AD-MCNN | 0.656 | **0.644**[a] | **0.647**[a] | **0.914** | **0.627** | **0.648** | **0.633** | **0.910** |

Bold fonts are utilized to indicate the top-performing algorithm for each metric

[a] Statistically significant over baselines

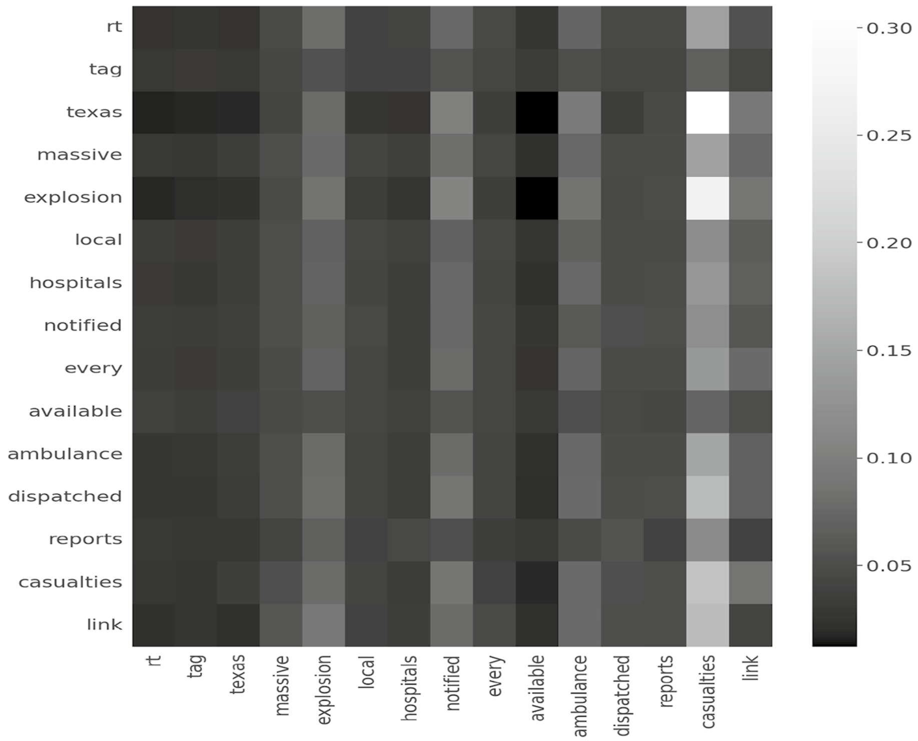


**Fig. 20** Visualisation of a Self-Attention head scores when applied to a sample tweet: *"RT @user: texas: massive explosion u/d - local hospitals notified. every available ambulance dispatched. reports of casualties. HTTP:..."*. The highest scores are observed in the combinations of location ("texas") and type of incident ("explosion") with a very important consequence "casualties". Note that each attention head is likely to focus on different combinations of words, providing attention to more parts of the tweet that may disclose additional information. In this example, one attention head may attend more to the casualties and others in the hospitals and ambulances

e.g., government and NGO, and (iii) *Information Type*, e.g., bombing, floods, explosion.

We experimented with a binary classification task on the *Informativeness* annotation level and also with a multi-class classification task using the *Information type* annotations. For the binary classification, all crisis events are concatenated into the event-related class and create an uneven class distribution between event and non-event posts. Therefore, we trained our model both for unbalanced and balanced settings, randomly oversampling the minority class in the latter setup. Moreover, we split the dataset with a 0.8–0.2 stratified split for training and testing, and then executed 10 runs for each experiment to calculate an average and standard deviation over different global network seeds. The experiments were run using a 12GB Nvidia GeForce RTX 2080 Ti.

For comparison purposes, we implemented two baseline architectures:

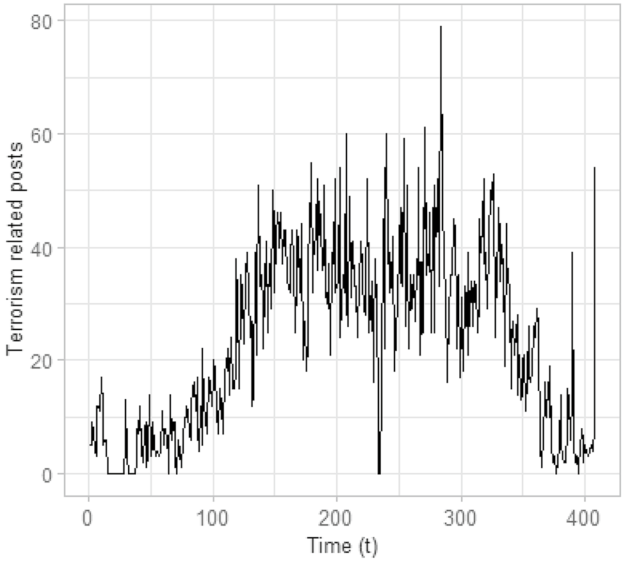


**Fig. 21** Time series of posts classified as terrorism-related

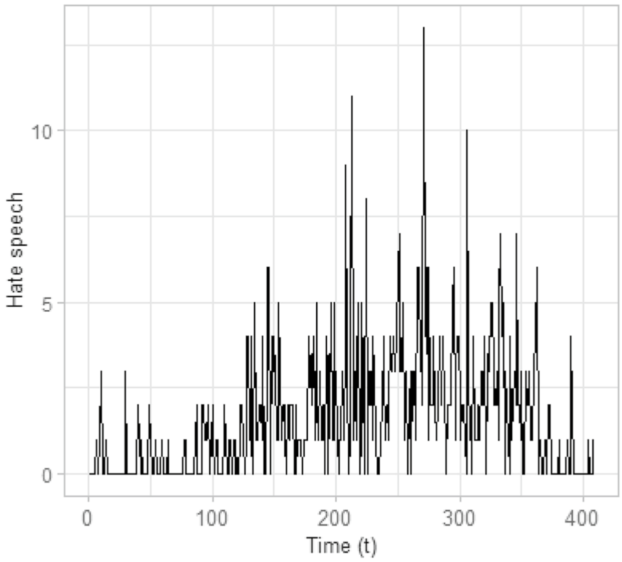


**Fig. 22** Time series of posts classified as terrorism-related or as containing hate speech

- **MCNN**: a Multi-channel CNN architecture [27], which has shown the best performance so far in the CrisisLexT26 dataset.
- **MCNN-MA**: a recently proposed architecture for the Sentiment Analysis task [82], which we adapted to our task with suitable hyper-parameterization. We consider this as a baseline, as it also uses multi-head attention after MCNN; in comparison, in our case, a full self-attention encoder is used before the MCNN to extract the most important features from the text.

For measuring the performance of each setup, standard evaluation metrics were used: Precision, Recall, F1-score, and Area Under the ROC Curve (AUC). The results are presented in Table 6. Binary classification results suggest that the proposed architectures exceed the baselines in F1 and AUC, both for balanced and unbalanced datasets, with AD-MCNN being undoubtedly the best performing model. This demonstrates that utilising the self-attention encoder as a denoiser, reinforcing important keywords, and reducing irrelevant ones improves detection performance, since it is not limited by the distance between words or the words between them, as an RNN is. On the contrary, the convolution window limits MCNN's feature extraction to surrounding words, hence employing attention after the CNNs (MCNN-MA) limits the capacity of the signal.

The overall findings for the multi-class classification exhibit behaviour that is comparable to that of the binary classification. The MCNN, however, performs better than previously. Although we observe similar performance on the balanced configuration, we contend that if more samples per class were provided (on average ≈2.8k original samples), the AD-MCNN's advantage over MCNN would grow.

**Table 7** Estimated change points for the two-dimensional time series along with the corresponding significance values

| Time | Date | p value |
|---|---|---|
| 77 | 23/02/2009 | 0.0033 |
| 119 | 06/04/2009 | 0.0033 |
| 148 | 05/05/2009 | 0.0033 |
| 216 | 12/07/2009 | 0.0033 |
| 237 | 02/08/2009 | 0.0432 |
| 346 | 19/11/2009 | 0.0133 |
| 373 | 16/12/2009 | 0.0498 |

Finally, our assertion that the attention mechanism serves as a denoiser for the text before it is passed on to other types of layers is depicted in an illustrative example in Fig. 20, where we apply self attention on a tweet and showcase that the scores of highly relevant words are matched up with higher scores, while non-important combinations exhibit low attention scores, leading to the claimed denoising behaviour. Therefore, important keyword combinations that essentially capture the essence of the event (*who, did-what, where*) aid in the faster and more accurate identification of important crisis incidents.

## Change Point Detection

This section showcases the applicability of the proposed terrorism-related change point detection framework tested

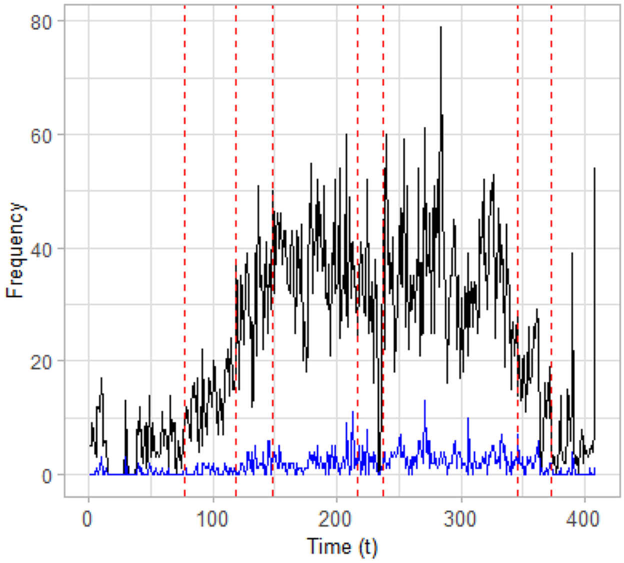


**Fig. 23** Application of CPD on the two-dimensional time series

on the publicly available *Ansar* dataset.[19] *Ansar* contains a collection of terrorism-related posts published in the Ansar AlJihad Network, a set of invitation-only jihadist forums in Arabic and English that are well-known among Western Jihadists [83]. The English part of the dataset, mentioned as *Ansar1*, includes 29,492 posts and spans the period between 8/12/2008–20/1/2010. The dataset contains some Arabic posts, which were removed resulting to 24,130 posts.

## Change Point Detection in Multivariate Time Series

This section focuses on the change point analysis of time series related to the extracted indicators of terrorism and hate speech relation aiming to estimate time locations of structural breaks in them that may indicate the occurrence of important events.

The CPD method applied in this work is described in [66] and takes into consideration both univariate and multivariate time series. Exploiting the outputs of the two classification models presented in the section “Change Point Detection in Terrorism-Related Online Content”, two time series are constructed and used as input to the CPD algorithm: (i) the time series of posts classified as terrorism-related, and (ii) those identified as containing hate speech. The posts are aggregated on a daily basis resulting in two time series of length $T = 408$ (days) which are presented in Figs. 21 and 22, respectively; regardless of the fact that the observed frequencies at the time series of terrorism-related posts are considerably higher, the two time series of interest seem to evolve in a similar manner.

The case of the multivariate time series enables to exploit possible correlations that may exist between the two constructed time series of terrorism- and hate-speech-related posts; therefore, we proceed with the combination of the two time series into a single, two-dimensional one $\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_T$, $T = 408$, $\mathbf{Z}_i = (z_{i,1}, z_{i,2})$, $i = 1, 2, \ldots, 408$, where the first entry of the observation vector $\mathbf{Z}_i$ is the frequency of the posts classified as terrorism-related and the second one denotes the frequency of the posts classified as containing hate speech. The concept behind combining terrorism-related posts with hate speech is that hate speech, in the sense of expressing aggressive behaviours, can be linked to terrorism and vice versa. The results of the two-dimensional CPD are presented in Table 7 and depicted graphically in Fig. 23.

Taking into consideration the time locations of the estimated change points, as well as the terrorist incidents that occurred during 2009,[20] which covers the main part of the *Ansar1* dataset, some conclusions could be derived. In general, it can be argued that the intensity of online activity appears to be closely related to the intensity of terrorism or crime occurrences. This is strengthened by the fact that a considerable number of terrorist incidents occurred (e.g., bomb explosion in Afghanistan on 25/03/2009 and suicide bombing in Pakistan on 27/03/2009) when either increasing trends are identified in online activity (i.e., 23/02/2009–06/04/2009 and 06/04/2009–05/05/2009) or the activity is stable at a high frequency level (i.e., 05/05/2009–12/07/2009). Especially, regarding the period where steady activity at high frequencies is identified, this characteristic may also be partially attributed to the previous period’s terrorist incidents, which caused an intense online activity that was maintained on account of the consequences of the attacks. Furthermore, it is concluded that the estimated change points related to growing trends correspond partially to time locations of terrorist occurrences. For example, this holds with the identified change points at $t = 77$ (23/02/2009) and $t = 119$ (06/04/2009), which both indicate the onset of periods with increasing trends; for more details about the interpretation of the outputs and the links to terrorism occurrences, please check [63].

[19] https://www.azsecure-data.org/dark-web-forums.html.

[20] A list of widely known terrorists attacks can be found for example in
(i) https://en.wikipedia.org/wiki/List_of_terrorist_incidents_in_2009.
(ii) https://www.dni.gov/nctc/index.html or in (iii) [84].

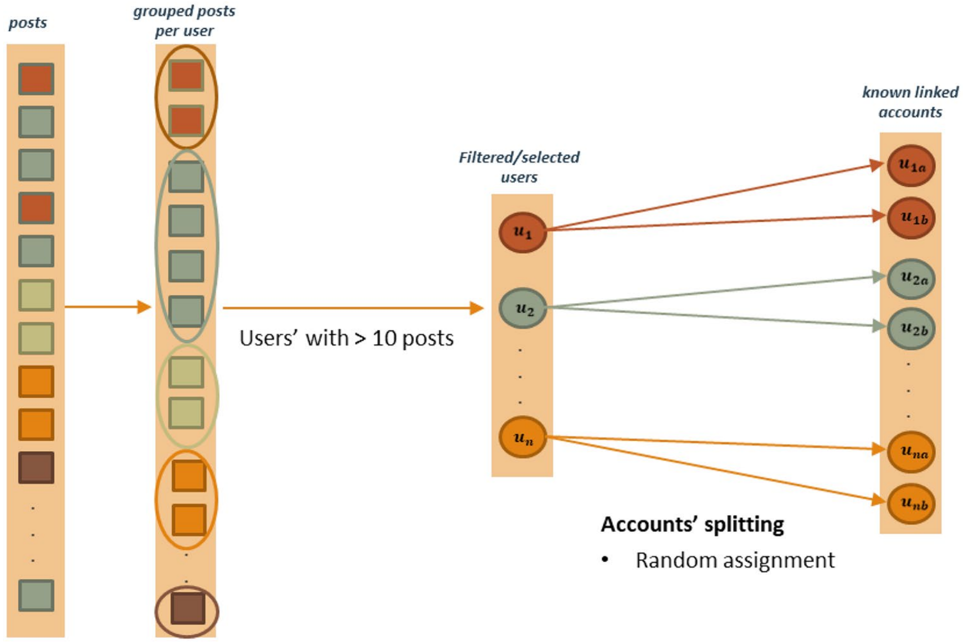


**Fig. 24** Creation of ground a truth dataset

**Table 8** Results of the User Identity Linkage module

| | Abusive case | | | | Terrorism case | | | |
|---|---|---|---|---|---|---|---|---|
| | Acc | AUC | Prec | Rec | Acc | AUC | Prec | Rec |
| Activity | 90.90 | 65.00 | 83.00 | 91.00 | 90.90 | 78.23 | 83.00 | 91.00 |
| Linguistic | 94.77 | 91.65 | 94.00 | 95.00 | 96.09 | 96.40 | 96.00 | 96.00 |
| Network | 90.86 | 81.41 | 83.00 | 91.00 | 91.00 | 83.26 | 92.00 | 91.00 |
| All | 95.90 | **96.05** | **96.00** | **96.00** | 96.18 | 97.99 | 96.00 | 96.00 |
| Activity+edits+sem | 90.90 | 68.77 | 83.00 | 91.00 | 93.22 | 93.52 | 93.00 | 93.00 |
| Linguistic+edits+sem | 94.22 | 90.82 | 94.00 | 94.00 | **96.68** | 97.24 | **97.00** | **97.00** |
| Network+edits+sem | 90.86 | 80.90 | 83.00 | 91.00 | 94.54 | 92.12 | 94.0 | 95.00 |
| All+edits+sem | **95.95** | 95.91 | **96.00** | **96.00** | 96.59 | **98.45** | **97.00** | **97.00** |

Bold fonts are utilized to indicate the top performing algorithm for each metric

## User Identity Linkage

To evaluate the user identity linkage framework, we focused on two types of delinquent behaviours and content, i.e., abusive and terrorism-related. In the abusive-related case, we use a dataset provided by [85] that was created for studying abusive activities on Twitter; it was collected between June and August 2016 and consists of 600k tweets in English and 312k users. Regarding the terrorism-related case, the data was collected using a set of terrorism-related Arabic keywords provided by Law Enforcement and domain experts; it spans from February 2017 to June 2018 and consists of 65k tweets and 35k users.

Since there are no ground truth datasets indicating which user accounts belong to the same natural person, we had to create them artificially. As depicted in Fig. 24, from the above-mentioned datasets we filter out all users with less than 10 posts, while then we randomly select a subset of $X$ (e.g., $X = 200$) user accounts by applying stratified random sampling.

Then, similar to other works [41, 42], we split the posts of each selected user account into two subsets and assign to each subset a different user id. For instance, user $u_i$ becomes $u_{ia}$ and $u_{ib}$ and the tweets of $u_i$ are split between $u_{ia}$ and $u_{ib}$ and therefore we consider $u_{ia}$ and $u_{ib}$ as being linked accounts. Thus, we come up with a dataset with the double number of user accounts and a set of known linked accounts. For instance, if in the data sampling step we had selected 200 user accounts, after the splitting process we result to 400 users. To split the tweets of the original accounts into linked users we randomly assigned an equal number of posts to each user.

In the end we end up with two sets of users, i.e., $A = \{u_{1a}, u_{2a}, \dots, u_{Xa}\}$ and $B = \{u_{1b}, u_{2b}, \dots, u_{Xb}\}$. Comparing each user $u_{ia}$ from set $A$ with each user $u_{ib}$ from set $B$ (where

**Table 9** Fine tuning the undersampling ratio

| Ratio | No. of Tweets | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| – | 9613 | 0.8185 | 0.6977 | 0.8678 | 0.7656 |
| 0.6 | 9125 | 0.8296 | 0.8342 | 0.6715 | 0.7331 |
| 0.7 | 8310 | 0.8291 | 0.7897 | 0.7675 | 0.7694 |
| 0.8 | 7699 | 0.8481 | 0.8036 | 0.8818 | **0.8364** |
| 0.9 | 7224 | 0.809 | 0.7332 | 0.9529 | 0.8223 |
| 1 | 6844 | 0.8262 | 0.8741 | 0.7771 | 0.8172 |

Bold fonts are utilized to indicate the top performing algorithm for each metric

**Table 10** Result with text pre-processing

| Pre-processing method | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| No pre-processing | 0.8169 | 0.8175 | 0.7770 | 0.7880 |
| Removed links, emojis | 0.8117 | 0.7429 | 0.9031 | 0.8098 |
| Removed links, hashtags, emojis | 0.8481 | 0.8036 | 0.8818 | **0.8364** |

Bold fonts are utilized to indicate the top performing algorithm for each metric

$i \neq j$), we result to overall $Y = X * (X - 1)$ user pairs (e.g., for $X = 200$, $Y = 39,800$), with each user pair in $Y$ corresponding to a non-linked account. For each dataset, we opt to retain a proportion of 10% of linked and 90% of non-linked accounts, since previous work (e.g., [86]) has indicated that about 10% of users within a dataset tend to exhibit bad behaviour.

As mentioned, features from three categories (i.e., Activity, Linguistic, and Network) are considered for user modelling, while user pairs are modelled based on the absolute difference (*abs*); in addition to the absolute difference, three similarities measures (i.e., cosine similarity, Euclidean and Manhattan distances) have also been estimated, the results of which can be found in [44]. Therefore, the following approaches are evaluated: $Activity_{abs}$, $Linguistic_{abs}$, $Network_{abs}$, $All_{abs}$. In addition, each pair of users is modelled using the edit distance (*edits*) and semantic similarities (*sem*).

We use Keras with Theano[21] for the deep learning models. In all cases, we use repeated (5 times) 10-fold cross validation, which is less variable than the ordinary 10-fold cross validation [87].

Table 8 depicts the results obtained for both the Abusive and Terrorism-related cases. When comparing the three feature categories, we observe that the linguistic features contribute more compared to the activity and network ones. In addition, we observe that in both cases we obtain the best performance when all features are considered. Especially, in the terrorism-related case, the overall performance (in terms of AUC) is further enhanced when all the features are considered in addition to the texts' similarities (i.e., *edits* and *sem*).

**Table 11** Results in terms of accuracy

| Method | Accuracy |
|---|---|
| STA/LTA | 0.7726 |
| Z-score | 0.8301 |
| KDE | 0.8986 |
| KDE+CD | **0.9589** |

Bold fonts are utilized to indicate the top performing algorithm for each metric

Detecting accounts of the same user poses several difficulties, since often users alternate their behavioural patterns in an effort to stay under the radar of social media platforms. Overall, the results showed that the developed *user identity linkage* method is able to effectively detect potentially linked accounts created in an effort to spread non-legitimate or even illegal activities. Finally, an important observation is that the developed model performs well in different languages (i.e., English and Arabic).

## Estimating Relevance to Fire Disaster

For the training and the evaluation of the architecture mentioned in the section "Relevance Estimation"), we trained our model on a dataset of 9,613 annotated Greek tweets referring to fires, from which the 3,422 (35,6%) where labeled as relevant and the 6,191 (64.4%) as irrelevant. Cross-validation was used for evaluating the performance of the text classifier. The procedure begins by randomly splitting the training dataset into 80% for the training set, 20% for the test set and 10% of the training set for the validation set.

The training set is the data used for learning and fitting the parameters to the machine learning model, and the validation set is used to provide an unbiased evaluation of the model fitted on the training dataset while tuning the model hyper-parameters. The validation set of data was used to provide an unbiased evaluation of a final model fitted on the training dataset.

Finally, the evaluation of the model on the test set and the respective prediction results were compared against the human-annotated tags with the following performance metrics: Accuracy, Precision, Recall, and F1-score (Table 9).

In the training stage, the number of epochs is a challenging choice for avoiding over-fitting and under-fitting. Over-fitting is the good performance on the training data and poor in new data, and under-fitting refers to a model that cannot perform well on the training data nor generalize the new data. Too many epochs can lead to over-fitting of the training dataset, whereas too few may result in an under-fit.

[21] https://keras.io/backend/.

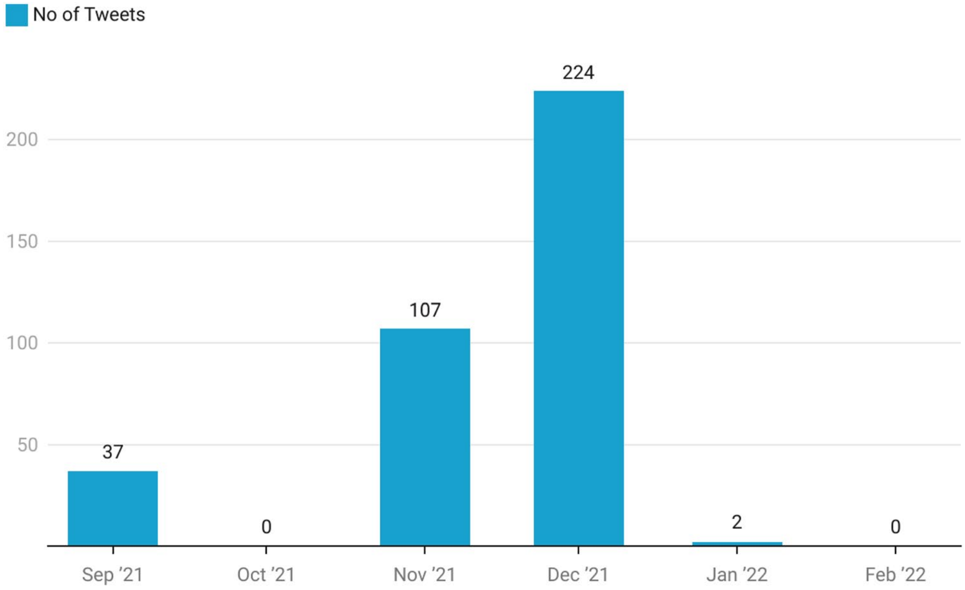


**Fig. 25** The number of tweets referred to air quality issues in Frankfurt city from September 2021 to February 2022

Therefore, we use an early stop callback to terminate the process early, before the validation loss increases and we save the model with the best validation F1-score.

Additionally, due to the imbalanced label distributions (relevant 35.6%, irrelevant 64.4%) the model can lead to poor performance by ignoring the minority class. To tackle this issue, the under-sampling method was applied by selecting and randomly removing some of the negative samples, in order to have the most balanced possible dataset with a class ratio closer to 1. In fact, a ratio equal to 1 corresponds to a balance of 50% for both relevant and irrelevant labels. Moreover, the best performing ratio was investigated and we concluded the fine-tuned ratio equal to 0.8, as can be seen in Table 9.

In the last evaluation stage, we considered the removal of the respective links, emoticons and hashtags of text, and we concluded the best performing model with an F1-score equal to 0.8364. Finally, the selection of the text pre-processing methods was investigated and concluded that the best approach is to remove hyperlinks, hashtags and emoticons (Table 10).

## Detection of Fire-Related Events

For the evaluation of the event detection methodology introduced in [72] and described in the section "Event Detection", we collected approximately 2 million tweets containing fire-related keywords in Spanish for the year of 2019. As ground truth we considered the 11 largest fires of 2019 that occurred in Spain, according the Copernicus Emergency Management Service.[22]

Our event detection methodology was compared with the afore-mentioned STA/LTA [78] and Z-score [77] for the above Twitter dataset. Both methods examine the fluctuation of the number of tweets published per day, unlike our method that investigates the temporal density of the posted tweets as well as the relationship between users. The performance of all methods was measured in accuracy and we took into consideration only the best performance for each threshold per method. In the results that are shown in Table 11 it is apparent that the best performing method by a significant margin is KDE+CD, i.e. the combination of Kernel Density Estimation and Community Detection, followed by plain KDE, Z-score, and STA/LTA, proving that the analysis of user communities has a respectably positive impact on the method's ability to detect real life events.

## Validation of Humans as Sensors for Air Quality

In this section we present the validation of social media data with actual measurements from air quality sensors. The validation dataset was collected from Twitter between September 6, 2021 and February 16, 2022, using German keywords

**Table 12** A sample of the search keywords

| List A | List B |
|---|---|
| Ground water | Strange smell |
| Spring | Contamination |
| Water | Pollution |
| Aquifer | Non drinkable |
| Domestic water | Turbid |
| Tap water | Strange smell |
| Etc | Etc |

[22] https://emergency.copernicus.eu/mapping/list-of-activations-rapid.

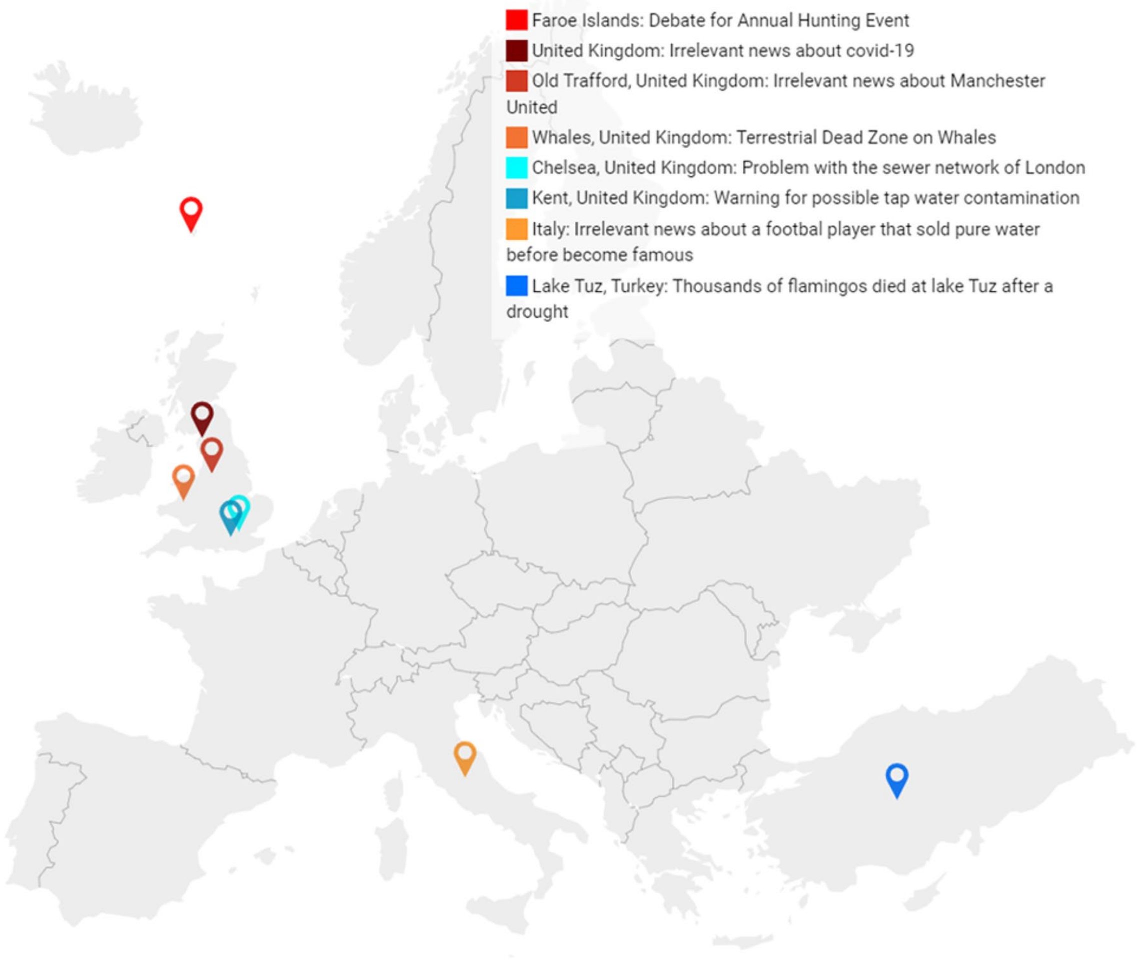

🟥 Faroe Islands: Debate for Annual Hunting Event
🟫 United Kingdom: Irrelevant news about covid-19
🟥 Old Trafford, United Kingdom: Irrelevant news about Manchester United
🟧 Whales, United Kingdom: Terrestrial Dead Zone on Whales
🟦 Chelsea, United Kingdom: Problem with the sewer network of London
🟦 Kent, United Kingdom: Warning for possible tap water contamination
🟧 Italy: Irrelevant news about a footbal player that sold pure water before become famous
🟦 Lake Tuz, Turkey: Thousands of flamingos died at lake Tuz after a drought

**Fig. 26** Events detected with Z-Score, STA/LTA, and DBSCAN

about air quality. The initial dataset collected consisted of 29,132 tweets, but the vast majority of tweets were lacking geoinformation. To combat this, we geotagged the Twitter posts with an NER implementation based on XML-RoBERTa (XML-R) language model [88] and with the assist of OpenStreetMap API for retrieving the exact coordinates. The final dataset comprised 2948 georeferenced tweets, approximately the 10% of the initial dataset.

Then, the geotagged tweets and the in-situ measurements from sensors were transformed into an RDF representation and were stored in a knowledge database. By applying SPARQL queries to the RDF knowledge graphs, we were able to fetch measurements of PM10 from air quality sensors in various German cities for the examined dates (September 2021 - February 2022) and compare them with the respective number of tweets per day.

One example of validation is that the number of tweets that discuss air quality issues in Frankfurt was increased during four days in December 2021, where the station "Kassel Funffensterstraße" measured a daily concentration of PM10 above the 50 μg/m$^3$ threshold (Fig. 25). The results show that humans indeed can be used as sensors for the detection of air quality, but more experiments with different datasets are needed for a robust correlation to be found.

## Detection of Water-Quality-Related Events

In order to evaluate our work introduced in [80] and presented in the section "Detection of Water Quality Incidents", we created a dataset of approximately 212,000 English tweets by searching for keywords that combine a water source (List A in Table 12) and a water issue (List B in Table 12), during one year (from August 1, 2020 to July 31, 2021). Additionally, circa 51,000 tweets of this dataset were automatically geotagged with NER, since one of the examined methods, i.e. DBSCAN, requires geoinformation.

On this dataset we applied the event detection techniques Z-score, LTA/STA, and DBSCAN in order to detect potential water quality incidents. Then, we tried to link these detected incidents into real life events and categorize them into relevant or not relevant water quality issues for the evaluation of the three models. In Fig. 26 we can discern an example of a relevant event, i.e. "Possible tap water contamination in Kent", and an irrelevant event, i.e. "Football player selling water".

It is of great interest that the results of the examined techniques were quite dissimilar. The LTA/STA method was the most precise with 6 relevant and 1 irrelevant event found. On the other hand, the DBSCAN method discovered the most events, out of which 8 were relevant and 4 irrelevant. Lastly, the Z-score method performed the worst, with only 2 relevant events and 3 irrelevant.

## Conclusions and Future Research Directions

In this paper we sought an answer to the question of how social media can be exploited and how the challenges that they pose in data mining can be overcome in order to provide useful data as input for applications. To this end, we explored content analysis, machine learning, network analysis, etc. in order to implement various techniques that can enhance the usage of social media data in the areas of architectural design, cultural applications, crime prediction, detection and prevention, disaster management, and water management.

For architecture design we developed a tool that inspires and supports architects by transforming social media textual and visual data into useful 3D semantic objects ready for use. The future goal is to extend the social media base, leveraging the wide amount of alternative sources that can be mined. For example, tips from the Foursquare platform can help text analysis provide further insights about architectural elements in landmarks. Connecting to new platforms, in combination with continuously collecting data from the already considered ones, can also lead to new datasets for training and validating multimodal analysis models.

For cultural applications we developed tools that collect thematic textual and visual content from social media platforms. The extracted data are used by experts to create repositories and immersive experiences which safeguard cultural heritage. Next steps are the extension of social media sources as well as the utilization of new methods for data mining and analysis. This will provide new valuable content that together with the present data collections, will enhance decision support and the algorithms training.

As a means of overcoming the background noise present in short social media messages (e.g. Twitter data), we developed three models aimed at improving detection capabilities for crisis-related events [56]. We postulated that by paying close attention to combined keywords, one could create a denoiser that would highlight salient features, so that each successive vector in the sequence would benefit from the additional information provided by the vectors representing closely related words. In order to test this hypothesis, we developed and analysed three attention-enhanced models that outperformed robust control baselines. We plan to conduct additional evaluations of our models in the future, particularly when more data is available, and to assess the influence of more current language models for word embeddings (such as [69, 89]), particularly multilingual ones [88].

Overall, taking into consideration that the time locations of the estimated change points are partially connected to the occurrence of terrorist incidents, it could be argued that the proposed change point detection framework serves as an alternative way for the identification of links between terrorism and activity in social media sources. Moreover, depending on the platform and the participants, the depiction of a more intense online activity related to terrorism or criminal activities in general may precede the occurrence of events, rendering the suggested framework to be identified also as an early warning tool. Since the focus of the proposed framework was on the offline change point detection methods, for future work we intend also to adopt online CPD algorithms aiming to detect the initialisation of radicalisation or criminal behaviours in (near) real time. Furthermore, additional indicators related for instance to sentiment analysis towards an event of interest could be extracted from online activity, gathering in this way more intelligence to enhance the input of the multivariate change point detection approach.

Due to the massive use of social media, deviant content is also created, such as abusive or even terrorist-related. In their efforts to bypass social media mitigation mechanisms and to further spread their ideas, ideals, or even their propaganda, users often create multiple accounts. Detecting multiple accounts of the same user is an especially challenging task due to the fact that individuals often change their behavioural patterns in order to refrain from being detected by social media platforms. For the developed User Identity Linkage framework to be as effective as possible, different types of features were considered, i.e. profile, content, and network based, to identify accounts that may belong to the same natural person. As future activities, additional social media platforms, such as Reddit, will be explored to examine the applicability of the developed framework to platforms of a somewhat different form. Moreover, additional attributes of user behaviour could be taken into account, capturing, for instance, expressions of sarcasm and irony, thus making the various behavioural patterns even more distinguishable.

It is a common occurrence for a tweet collected based on certain search keywords to have a completely different context from what is collected for. To solve this problem we proposed a machine learning model that uses the text features, extracted from tweets in Greek language, to identify which posts are indeed relevant in the domain of fire disasters. Future developments will include the extension of the model to support other languages as well as the exploration of dataset augmentation techniques for enhancing the existing training sets that are not adequate for all languages.

For detecting disaster events we proposed a methodology [72] that uses two modalities: a Kernel Density Estimation algorithm that calculates the Density Score and a Community Detection algorithm that computes the Modularity Score. The two scores are fused and the model estimates if there is an event. For its evaluation, we compared KDE+CD with well known techniques such as Z-score and LTA/STA and found out that the performance of our approach is superior. Future actions could be the extension of the method to use other modalities and the evaluation for one-hour time frames instead of one-day time periods.

For the detection of creeping crisis incidents we presented a framework that recognizes air quality issues [74]. We collected Twitter data based on German keywords about air quality as well as in-situ measurements from air quality sensors. Next, we georeferenced the tweets with a state-of-the-art NER implementation and transformed both data into RDF semantic representation in order to use SPARQL language for querying and linking sensor observations with crowdsourcing information. Further efforts will include exploring more open databases, so as to find more correlations, and providing more analytics in our results. Finally, we could enhance the SPARQL queries with GeoSPARQL functionalities in order to further analyze the geospatial data (e.g. distance and overlap between locations).

In addition, we examined whether it is feasible to discover water quality related events through social media [80] by comparing event detection techniques, such as Z-score, STA/LTA and DBSCAN. We retrieved circa 200 thousand tweets posted during a year and used them for evaluation. The results showed significant differences between the above techniques, thus a future goal will be to find the balance between increasing the detection rate of relevant events and decreasing the noise.

In conclusion, the results of this multitude of presented methodologies showcased that social media data can indeed offer a meaningful added value to the five domains that we explored in this paper, as auxiliary means for real-events assessment, decision-making, prevention, and more. A future objective will be to implement an even larger number of applications to different domains and various scenarios as well as to use more sources of crowd-generated information.

**Funding** Open access funding provided by HEAL-Link Greece. This work has been supported by the TREEADS, INFINITY, WQeMS and CHROMATA projects. TREEADS, INFINITY, and WQeMS have received funding from the European Union's Horizon 2020 Research & Innovation Programme under Grant Agreements No. 101036926, No. 883293 and No. 101004157 respectively. CHROMATA has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH-CREATE-INNOVATE (project code: T2EDK-01856). This publication reflects only the authors' view and the Research Executive Office (REA) and the European Commission are not responsible for any use that may be made of the information it contains.

**Data availability** The social media post IDs from X (previously Twitter) are available upon request. Their acquisition from X are subject to the policy of X and the selected plan.

## Declarations

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Ethical Approval** Not applicable.

**Consent to Participate** Not applicable.

**Consent for Publication** Not applicable.

## References

1. Carr CT, Hayes RA. Social media: defining, developing, and divining. Atl J Commun. 2015;23(1):46–65.
2. Reuter C, Hughes AL, Kaufhold M-A. Social media in crisis management: an evaluation and analysis of crisis informatics research. Int J Hum-Comput Interact. 2018;34(4):280–94.
3. Middleton SE, Middleton L, Modafferi S. Real-time crisis mapping of natural disasters using social media. IEEE Intell Syst. 2013;29(2):9–17.
4. Said N, Ahmad K, Riegler M, Pogorelov K, Hassan L, Ahmad N, Conci N. Natural disasters detection in social media and satellite imagery: a survey. Multimed Tools Appl. 2019;78(22):31267–302.
5. Komito L. Social media and migration: virtual community 2.0. J Am Soc Inf Sci Technol. 2011;62(6):1075–86.
6. Dekker R, Engbersen G, Klaver J, Vonk H. Smart refugees: how Syrian asylum migrants use social media information in migration decision-making. Soc Media Soc. 2018;4(1):2056305118764439.
7. Zeng B, Gerritsen R. What do we know about social media in tourism? A review. Tour Manag Perspect. 2014;10:27–36.
8. Vo T, Sharma R, Kumar R, Son LH, Pham BT, Tien Bui D, Priyadarshini I, Sarkar M, Le T. Crime rate detection using social media of different crime locations and twitter part-of-speech tagger with brown clustering. J Intell Fuzzy Syst. 2020;38(4):4287–99.

9. Song Y, Zhang B. Using social media data in understanding site-scale landscape architecture design: taking Seattle freeway park as an example. Landsc Res. 2020;45(5):627–48.
10. Kim HJ, Chae BK, Park SB. Exploring public space through social media: an exploratory case study on the high line New York city. Urban Des Int. 2018;23(2):69–85.
11. Valls F, Redondo E, Fonseca D, Torres-Kompen R, Villagrasa S, Martí N. Urban data and urban design: a data mining approach to architecture education. Telemat Inform. 2018;35(4):1039–52. https://doi.org/10.1016/j.tele.2017.09.015.
12. Heverin T, Zach, L. Microblogging for crisis communication: examination of twitter use in response to a 2009 violent crisis in the Seattle-Tacoma, Washington, area. In: ISCRAM 2010.
13. LDC: ACE 2005 Multilingual Training Corpus 2005. https://catalog.ldc.upenn.edu/LDC2006T06.
14. Hong Y, Zhang J, Ma B, Yao J, Zhou G, Zhu Q. Using cross-entity inference to improve event extraction. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies; 2011. pp 1127–1136.
15. Li Q, Ji H, Huang L. Joint event extraction via structured prediction with global features. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); 2013. pp. 73–82.
16. Li Q, Ji H, Hong Y, Li S. Constructing information networks using one single model. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014. pp. 1846–1851.
17. Chen Y, Xu L, Liu K, Zeng D, Zhao J. Event extraction via dynamic multi-pooling convolutional neural networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers); 2015. pp. 167–176.
18. Nguyen TH, Grishman R. Event detection and domain adaptation with convolutional neural networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 2015. pp. 365–371.
19. Nguyen TH, Cho K, Grishman R. Joint event extraction via recurrent neural networks. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016. pp. 300–309.
20. Nguyen T, Grishman R. Graph convolutional networks with argument-aware pooling for event detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2018. p. 32.
21. Yan H, Jin X, Meng X, Guo J, Cheng X. Event detection with multi-order graph convolution and aggregated attention. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019. pp. 5770–5774.
22. Liu S, Chen Y, Liu K, Zhao J. Exploiting argument information to improve event detection via supervised attention mechanisms. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2017. pp. 1789–1798.
23. Olteanu A, Vieweg S, Castillo C. What to expect when the unexpected happens: Social media communications across crises. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, 2015.
24. Caragea C, Silvescu A, Tapia AH. Identifying informative messages in disaster events using convolutional neural networks. In: International Conference on Information Systems for Crisis Response and Management, 2016. pp. 137–147.
25. Nguyen D, Al Mannai KA, Joty S, Sajjad H, Imran M, Mitra P. Robust classification of crisis-related data on social networks using convolutional neural networks. In: Proceedings of the International AAAI Conference on Web and Social Media, 2017. p. 11.
26. Burel G, Saif H, Alani H. Semantic wide and deep learning for detecting crisis-information categories on social media. In: International Semantic Web Conference. Springer, 2017. pp. 138–155.
27. Burel G, Alani H. Crisis event extraction service (crees)-automatic detection and classification of crisis-related content on social media 2018.
28. Kabir MY, Madria S. A deep learning approach for tweet classification and rescue scheduling for effective disaster management. In: Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2019. pp. 269–278.
29. Asongu SA, Orim S-MI, Nting RT. Terrorism and social media: global evidence. J Glob Inf Technol Manag. 2019;22(3):208–28.
30. Liu S, Yamada M, Collier N, Sugiyama M. Change-point detection in time-series data by relative density-ratio estimation. Neural Netw. 2013;43:72–83.
31. Wang Y, Goutte C. Detecting changes in twitter streams using temporal clusters of hashtags. In: Proceedings of the Events and Stories in the News Workshop, 2017. pp. 10–14.
32. Tasoulis SK, Vrahatis AG, Georgakopoulos SV, Plagianakos VP. Real time sentiment change detection of twitter data streams. arXiv preprint 2018. arXiv:1804.00482
33. Everton SF, Cunningham D. Detecting significant changes in dark networks. Behav Sci Terrorism Polit Aggress. 2013;5(2):94–114.
34. Tickle S, Eckley I, Fearnhead P. A computationally efficient, high-dimensional multiple changepoint procedure with application to global terrorism incidence. arXiv preprint 2020. arXiv:2011.03599
35. Tsikerdekis M, Zeadally S. Multiple account identity deception detection in social media using nonverbal behavior. IEEE Trans Inf Forensics Secur. 2014;9(8):1311–21.
36. Homeland Security: The Role of Identity Resolution in Criminal Investigations 2018. https://www.netowl.com/role-identity-resolution-criminal-investigations.
37. Riederer C, Kim Y, Chaintreau A, Korula N, Lattanzi S. Linking users across domains with location data: Theory and validation. In: WWW, 2016; pp. 707–719. IW3C2.
38. Liu L, Li X, Cheung W, Liao L. Structural representation learning for user alignment across social networks. In: IEEE Transactions on Knowledge and Data Engineering, 2019.
39. Liu S, Wang S, Zhu F, Zhang J, Krishnan R. Hydra: Large-scale social identity linkage via heterogeneous behavior modeling. In: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, 2014. pp. 51–62.
40. Srivastava DK, Roychoudhury B, Samalia HV. Importance of user's profile attributes in identity matching across multiple online social networking sites. In: 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE; 2018. pp. 14–15.
41. Pennekamp J, Henze M, Hohlfeld O, Panchenko A. Hi doppelgänger: Towards detecting manipulation in news comments. In: Companion Proceedings of The 2019 World Wide Web Conference. ACM; 2019. pp. 197–205.
42. Johansson F, Kaati L, Shrestha A. Timeprints for identifying social media users with multiple aliases. Secur Inform. 2015;4(1):7.
43. Mishra I, Dongre S, Kanwar Y, Prakash J. Detecting users with multiple aliases on twitter. In: 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE; 2018. pp. 560–563.
44. Chatzakou D, Soler-Company J, Tsikrika T, Wanner L, Vrochidis S, Kompatsiaris I. User identity linkage in social media using

linguistic and social interaction features. In: 12th ACM Conference on Web Science. 2020; pp. 295–304.

45. Mahiuddin M. Real time sentiment analysis and opinion mining on refugee crisis. In: 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), 2019. pp. 699–705. https://doi.org/10.1109/ICAEE48663.2019.8975462.
46. Ginzarly M, Pereira Roders A, Teller J. Mapping historic urban landscape values through social media. J Cult Herit. 2019;36:1–11. https://doi.org/10.1016/j.culher.2018.10.002.
47. Nummi P. Crowdsourcing local knowledge with ppgis and social media for urban planning to reveal intangible cultural heritage. Urban Plan. 2018. https://doi.org/10.17645/up.v3i1.1266.
48. Sherchan W, Pervin S, Butler CJ, Lai JC, Ghahremanlou L, Han B. Harnessing twitter and Instagram for disaster management. IBM J Res Dev. 2017;61(6):1–8.
49. Chatfield AT, Brajawidagda U. Twitter early tsunami warning system: A case study in indonesia's natural disaster management. In: 2013 46th Hawaii International Conference on System Sciences. IEEE; 2013. pp. 2050–2060.
50. Aulov O, Halem M. Human sensor networks for improved modeling of natural disasters. Proc IEEE. 2012;100(10):2812–23.
51. Lambert LH, Bir C. Evaluating water quality using social media and federal agency data. J Water Health. 2021;19(6):959–74.
52. Zheng H, Hong Y, Long D, Jing H. Monitoring surface water quality using social media in the context of citizen science. Hydrol Earth Syst Sci. 2017;21(2):949–61.
53. Wickstrom AE, Specht AR. Tweeting with authority: Identifying influential participants in agriculture-related water quality twitter conversations. J Appl Commun. 2016;100(4):45–55.
54. Stathopoulos EA, Shvets A, Carlini R, Diplaris S, Vrochidis S, Wanner L, Kompatsiaris I. Social media and web sensing on interior and urban design. In: 2022 IEEE Symposium on Computers and Communications (ISCC), 2022. pp. 1–6. https://doi.org/10.1109/ISCC55528.2022.9912803
55. Kim, Y. Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1746–1751. Association for Computational Linguistics, Doha, Qatar 2014. https://doi.org/10.3115/v1/D14-1181. https://www.aclweb.org/anthology/D14-1181.
56. Kyriakidis P, Chatzakou D, Tsikrika T, Vrochidis S, Kompatsiaris I. Leveraging transformer self attention encoder for crisis event detection in short texts. In: European Conference on Information Retrieval. Springer; 2022. pp. 163–171.
57. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017. pp. 6000–6010.
58. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp. 770–778.
59. Ba JL, Kiros JR, Hinton GE. Layer normalization. Stat. 2016;1050:21.
60. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. J Mach Learn Res. 2011;12:2493–537.
61. Aue A, Horváth L. Structural breaks in time series. J Time Ser Anal. 2013;34(1):1–16.
62. Truong C, Oudre L, Vayatis N. Selective review of offline change point detection methods. Signal Process. 2020;167: 107299.
63. Theodosiadou O, Pantelidou K, Bastas N, Chatzakou D, Tsikrika T, Vrochidis S, Kompatsiaris I. Change point detection in terrorism-related online content using deep learning derived indicators. Information. 2021;12(7):274.
64. Davidson T, Warmsley D, Macy M, Weber I. Automated Hate Speech Detection and the Problem of Offensive Language, 2017.
65. de Gibert O, Perez N, García-Pablos A, Cuadros M. Hate speech dataset from a white supremacy forum. In: Proceedings of the 2nd Workshop on Abusive Language Online (ALW2), pp. 11–20. Association for Computational Linguistics, Brussels, Belgium 2018. https://doi.org/10.18653/v1/W18-5102. https://www.aclweb.org/anthology/W18-5102.
66. Matteson DS, James NA. A nonparametric approach for multiple change point analysis of multivariate data. J Am Stat Assoc. 2014;109(505):334–45.
67. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space, 2013. CoRR arXiv:1301.3781.
68. Pistola T, Diplaris S, Stentoumis C, Stathopoulos EA, Loupas G, Mandilaras T, Kalantzis G, Kalisperakis I, Tellios A, Zavraka D, Koulali P, Kriezi V, Vraka V, Venieri F, Bacalis S, Vrochidis S, Kompatsiaris I. Creating immersive experiences based on intangible cultural heritage. In: 2021 IEEE International Conference on Intelligent Reality (ICIR), 2021. pp. 17–24. https://doi.org/10.1109/ICIR51845.2021.00012.
69. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint 2018. arXiv:1810.04805.
70. Efron M, Lin J, He J, De Vries A. Temporal feedback for tweet search with non-parametric density estimation. In: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, 2014. pp. 33–42.
71. De Meo P, Ferrara E, Fiumara G, Provetti A. Generalized louvain method for community detection in large networks. In: 2011 11th International Conference on Intelligent Systems Design and Applications. IEEE; 2011. pp. 88–93.
72. Papadimos T, Pantelidis N, Andreadis S, Bozas A, Vrochidis S, Kompatsiaris I. Real-time alert framework for fire incidents using multimodal event detection on social media streams. In: 19th International Conference on Information Systems for Crisis Response and Managements, 2022.
73. Boin A, Ekengren M, Rhinard M. Hiding in plain sight: conceptualizing the creeping crisis. Risk, Hazards Crisis Public Policy. 2020;11(2):116–38.
74. Andreadis S, Mavropoulos T, Pantelidis N, Vrochidis S, Elias M, Papadopoulos C, Gialampoukidis I, Kompatsiaris I. Sparql querying for validating the usage of automatically georeferenced social media data as human sensors for air quality. In: 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP). IEEE; 2022. pp. 1–5.
75. Wang Z, Ye X. Social media analytics for natural disaster management. Int J Geogr Inf Sci. 2018;32(1):49–72.
76. Luna S, Pennock MJ. Social media applications and emergency management: a literature review and research agenda. Int J Disaster Risk Reduct. 2018;28:565–77.
77. Shiffler RE. Maximum z scores and outliers. Am Stat. 1988;42(1):79–80.
78. Allen RV. Automatic earthquake recognition and timing from single traces. Bull Seismol Soc America. 1978;68(5):1521–32.
79. Ester M, Kriegel H-P, Sander J, Xu X, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. KDD. 1996;96:226–31.
80. Andreadis S, Pantelidis N, Gialampoukidis I, Vrochidis S, Kompatsiaris I. Water quality issues: can we detect a creeping crisis with social media data? In: Proceedings of Symposium on Computers and Communications (ISCC). IEEE; 2022.
81. Dutta A, Zisserman A. The VIA annotation software for images, audio and video. In: Proceedings of the 27th ACM International Conference on Multimedia. MM '19. ACM, New York, NY, USA 2019. https://doi.org/10.1145/3343031.3350535.

82. Feng Y, Cheng Y. Short text sentiment analysis based on multichannel cnn with multi-head attention mechanism. IEEE Access. 2021;9:19854–63.
83. Scanlon JR, Gerber MS. Automatic detection of cyber-recruitment by violent extremists. Secur Inform. 2014. https://doi.org/10.1186/s13388-014-0005-5.
84. Burke RA. Counter-terrorism for emergency responders. 3rd ed. Boca Raton, Florida: CRC Press; 2017.
85. Chatzakou D, Kourtellis N, Blackburn J, De Cristofaro E, Stringhini G, Vakali A. Measuring #gamergate: A tale of hate, sexism, and bullying. In: Proceedings of the 26th International Conference on World Wide Web Companion. IW3C2; 2017. pp. 1285–1290.
86. Kayes I, Kourtellis N, Quercia D, Iamnitchi A, Bonchi F. The social world of content abusers in community question answering. In: Proceedings of the 24th International Conference on World Wide Web. IW3C2; 2015. pp. 570–580.
87. Kim J-H. Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap. Comput Stat Data Anal. 2009;53(11):3735–45.
88. Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, Grave É, Ott M, Zettlemoyer L, Stoyanov V. Unsupervised cross-lingual representation learning at scale. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020. pp. 8440–8451.
89. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V. Roberta: a robustly optimized bert pretraining approach. arXiv preprint 2019. arXiv:1907.11692.