RESEARCH ARTICLE-COMPUTER ENGINEERING AND COMPUTER SCIENCE

Automating Global Threat-Maps Generation via Advancements of News Sensors and AI

Fahim K. Sufi[1] · Musleh Alsulami[2] · Adnan Gutub[3]

Abstract

Negative events are prevalent all over the globe round the clock. People demonstrate psychological affinity to negative events, and they incline to stay away from troubled locations. This paper proposes an automated geospatial imagery application that would allow a user to remotely extract knowledge of troubled locations. The autonomous application uses thousands of connected news sensors to obtain real-time news pertaining to all global troubles. From the captured news, the proposed application uses artificial intelligence-based services and algorithms like sentiment analysis, entity detection, geolocation decoder, news fidelity analysis, and decomposition tree analysis to reconstruct global threat maps representing troubled locations interactively. The fully deployed system was evaluated for full three months of summer 2021, during which the autonomous system processed above 22 k news from 2397 connected news sources involving BBC, CNN, NY Times, Government websites of 192 countries, and all possible major social media sites. The study revealed 11,668 troubled locations classified successfully with outstanding precision, recall, and F1-score, all evaluated in ubiquitous environment covering mobile, tablet, desktop, and cloud platforms. The system generated interesting global threat maps for robust scenario set of 3.71×10^{29}, to be reported as original fully autonomous remote sensing application of this kind. The research discloses attractive news and global threat-maps with trusted overall classification accuracy.

1 Introduction

Previous research has successfully used geospatial mapping to better understand the emotions of theme park visitors [1], sentiment on wireless services [2], occurrence of natural disasters (like flood) [3], and even sentiment on political agendas analyzing Twitter messages [4]. However, these studies in [1, 2, 4] used geotagged location information found within a small fraction of the tweet messages. According to [3], tweets with geotagged location could be found only within 1% of the vast tweet messages. This means studies in [1, 2, 4] failed to utilize 99% of tweet messages without location information (i.e., non-geotagged tweets). Other weakness of processing geotagged tweets is that with the mobility of people, tweets can hold on to the previous address of the user. Additionally, in case the location services are disabled by the user, research from tweet messages could result in flaw in analysis.

Because of the drawback of using geotagged tweet (i.e., tweets with location information), researchers in [3] extracted possible location information from the tweet text by conducting a lookup search against a locally maintained location table. Even though using lookup against the tweet text with location table means utilizing the almost 100% of the tweet messages (as opposed to only 1% of geotagged tweets), it is only suitable for a selected region of interest. If the method demonstrated in [3] is to be implemented globally across all the regions, then the requirement of managing a vast location table with all possible locations in the world becomes unsustainable for a small- or medium-sized organization. Hence, a recent research in the extraction of location

✉ Fahim K. Sufi
  research@fahimsufi.com

✉ Musleh Alsulami
  mhsulami@uqu.edu.sa

1 Federal Government, Melbourne, VIC 3000, Australia

2 Information Systems Department, Umm Al-Qura University (UQU), Makkah 24382, Saudi Arabia

3 Computer Engineering Department, Umm Al-Qura University (UQU), Makkah 24382, Saudi Arabia

information from social media has been using Named Entity Recognition (NER) [5, 6], a more sustainable approach.

With the advent of internet, high-performance cloud infrastructure, and ubiquitous sensor networks, negative news travel faster than ever. As a negative incident occurs at a location, people take pictures and propagate the news within social media, and then, major news agencies could relay that news highlighting the location of incident. Since people are more attentive and attracted to negative news and events [7], news agencies diligently report almost all negative news events around the globe [8, 9]. Therefore, mainstream social media platforms (including Facebook, Snapchat, Telegram, Twitter, and Instagram), online news agencies, and trusted websites belonging to government and non-government entities can potentially become sensors for sensing troubled global locations.

This paper proposes an architecture of a system that uses application programming interfaces (API) for all the mainstream social media services like Facebook, Snapchat, Twitter, and Instagram, and web scraping technologies for online news agencies that senses all the major sources for major events. As seen from Fig. 1, after sensing the news, threat map is generated from the news using artificial intelligence (AI)-based services that include sentiment detection, NER, decomposition analysis, and an algorithm that calculates the trustworthiness of the news. Once the threat maps of the troubled locations are reconstructed with news data fusion, an intelligence officer or a strategic decision maker can interactively perform multi-dimensional analysis of the threat maps from his or her mobile devices.

The proposed architecture was implemented in fully managed cloud environment using the most recent version of AI workspace. The system was deployed in cloud, desktop, tablet, and mobile environment (both iOS and Android) and thoroughly evaluated for the threat-map generation functionalities. From June 2, 2021, to September 1, 2021, the deployed remote sensing application automatically sensed global events from 2397 media sources (e.g., government websites of 192 countries, major news agencies like BBC, CNN, NY Times, etc., and all major social media sites like Facebook, Instagram, Twitter, etc.). During this timeframe, 22,425 news were monitored. Out of all these news reports, 15,902 news were highlighted automatically as high-fidelity news with different combinations of news sentiments. Furthermore, 11,668 locations were correctly identified with a location classification accuracy of 0.992. The classification accuracy of the entire system harnessed by applying sentiment detection, news fidelity detection, and geolocation decoding was found to be of 0.992 precision, 0.993 recall, 0.993 F1-score, and 0.985 accuracy. According to the literature, this paper reports the first study to construct worldwide threat maps of troubled locations covering $3.71 X 10^{29}$ scenarios obtained from the

widest range of media sources with a very high level of classification accuracy. A fully managed real-time application of the proposed solution would allow government decision makers and intelligence officers to obtain location-based global intelligence in an interactive manner. In summary, the proposed solution has made the following contributions:

1) By using NER, the proposed approach automatically extracted location information from the global news sources and produced global threat maps of troubled locations. Most of the existing research demonstrated in [1–4, 8, 10–18] did not demonstrate automated location extraction mechanisms with AI-based entity recognitions.

2) Unlike the existing research works that analyzed news reports only in English language (as demonstrated in [1–4, 8, 10–20]), the presented approach can detect and understand news reports in many supported languages (e.g., Hindi, Bengali, German, Arabic, etc.).

3) Existing researchers in [1–4, 8, 10–20] only used one or two news sources. By harnessing news sensor technology described within this paper, the proposed approach aggregated news from 2397 different news sources.

4) Lastly, this is the first study reported in the literature to perform root cause analysis on global news reports using decomposition tree analysis.

2 Background Literature

Researchers in [1] provided a unique perspective in combining dictionary-based sentiment analysis with location information in visualizing theme park visitors' emotions in map. This research used Circumplex Model with two orthogonal axes of valence and arousal to identify 28 emotional expressions and differentiate emotional types by two axes or four quadrants (shown in Fig. 1, Table 1 in [1]). Sentiment analysis of tweet texts originating from Disneyland theme park was performed by dictionary search or lookup against 28 of these emotion types (considering all the variance of these words of emotions). At the end, four quadrants of emotions (pleasure level vs. arousal level) were obtained with tweet locations to understand the emotions of theme park visitors around different parts of the theme park. However, in [1], NER was not performed within this research. Instead, the researchers only used geotagged information from Twitter (meaning that only a fraction of the vast tweet messages was utilized).

In [2], multiple keywords related to wireless signal were used to filter out relevant geotagged tweets for performing
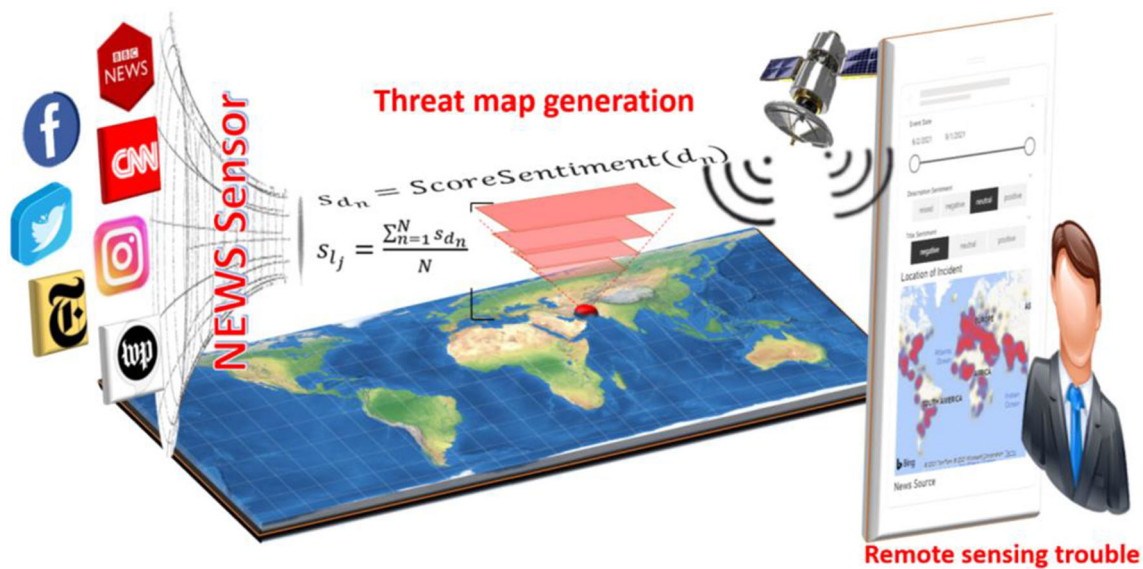
Fig. 1 Remote sensing trouble with 2397 deployed news sensors and threat-map generation process

Table 1 Existing study in sentiment analysis with media sources and algorithm used

References	Event source used	Year	Usage of named entity detection
[3]	Twitter	2014	No, Location extraction with Lookup
[1]	Twitter	2020	No, Twitter Geotagged Info used
[2]	Twitter	2019	No, Twitter Geotagged Info Used
[17]	BBC News Database	2017	No
[18]	COAE2014 (BBC Dataset)	2017	No
[13]	Reviews of Electronic product from Twitter	2018	No
[15]	Amazon Customer Review Data	2018	No
[16]	Hotel Reviews from OpinRank	2018	No
[10]	Facebook	2018	No
[11]	Instagram	2018	No
[20]	Social Media (unspecified)	2013	Yes
[19]	French patient forum	2018	Yes
[14]	Twitter	2017	No
[12]	Twitter	2019	No
[4]	Twitter	2021	No
[8]	Summary of World Broadcast, New York Times	2011	No

sentiment analysis using a range of machine learning algorithms.

The study in [3] is one of the early works presenting the applicability of a real-time crisis mapping of natural disasters using Twitter data. Unlike the previous works, this work uses geo-parsing to extract location information from the tweet text rather than using the geotag information from the tweet text. Extracting location information from the tweet text allows accessibility and utilization of 100% of the tweet messages rather than only 1% of geotagged tweets [3]. However, the presented system in [3] had to maintain a large set of location for the region of interest, as the solution do not perform location extraction with NER and rather perform lookup operation against locally maintained location table to extract location information from Twitter data. While this is suitable for a limited number of region of interest, this solution is resource-intensive and unsustainable to continuously update and maintain all location changes around the globe. Hence, this solution in [3] is not practical for global monitoring on all situations including all types of disasters.

The existing body of knowledge in worldwide news sentiment demonstrates that general sentiments in news published by the news agencies are becoming more and more negative [8, 9]. This is mainly because public are attracted toward negative events and the media agencies are supplying the demand of more negative news toward their consumers [7]. The analysis of worldwide media with sentiment detection and opinion mining is tremendously important because it can divulge the changes in tone of negativity toward another foreign entity, group, or locations [8]. More importantly, by scrutinizing the steady alteration in news sentiments, it is also feasible to envisage probable events in the future [8].

Given the importance of deeply analyzing the sentiments of worldwide events, most of the existing studies only obtained their media message from one or two media sources as seen from Table 1. The existing literature mainly depended on single source of information like Facebook [10], Instagram [11], Twitter [4, 12–14], Amazon review [15], and online blogs [16, 19] for conducting sentiment analysis research. Very few sources also used websites of news agencies like New York Times, BBC, Summary of World Broadcast, etc. [8, 17, 18]. Other than very few studies like [19, 20], majority of the existing studies did not utilize entity detection methods or NER for performing multi-dimensional analysis. In this section, we will briefly summarize the existing research on sentiment analysis, NER, and decomposition tree analysis.

Sentiment analysis technique uses natural language processing to find out if the particular text is positive, negative, or neutral. Sentiment analysis has been traditionally used for movie reviews or obtaining customer feedbacks. A research on sentiment analysis commenced in 2002 with two publications: [21, 22]. [21] described supervised learning corpus-based machine classifier, and on the other hand, [22] described unsupervised classifier based on linguistic analysis. There have been a series of publications on sentiment analysis, opinion mining, and emotion analysis [4, 12, 19–27].

NER or entity detection uses natural language processing technologies for extracting information from unstructured texts into predefined entity categories. These entity categories could be location names, landmarks, person names, organizations, date/time expressions, codes, numeric quantities, currency values, percentages, etc. Previous studies in entity detection have been presented in [20, 28]. The research in [28] extracted limited set of entities like "Pharmacologic Substance," "Disease or Syndrome," and "Sign or Symptom." Research in [20] extracted *Opioids, Drug Abuse—Cannabinoids*, *Sedatives*, *Buprenorphine* as well as *Stimulants*. Moreover, [19] extracted *methylphenidate* from five French patient web-forums.

Decomposition tree visualization is a graphical tool based on Microsoft Power BI. This visualization assists in guided root cause analysis as well as ad hoc exploration for data. Decomposition analysis supports drilling down into one or more factors by either high values or low values as shown in our recent study on global landslides [29, 30].

As seen from Table 1, Twitter was used in [1–4, 12–14], Facebook was used in [10], and Instagram was used in [11]. Other studies used online news, online forums, review data, etc., as in the case of [8, 15–20]. In terms of machine learning algorithms, support vector machine (SVM) was used in [2, 4, 10, 14, 15], naïve Bayes (NB) was used in [2, 15, 16], and K-nearest neighbor with decision tree was used in [10]. However, none of these previous studies in [1–4, 10–20] used

decomposition tree analysis method. Moreover, other than [19] and [20], none of these studies used AI-based entity recognition. In this study, we use entity detection and decomposition tree analysis to generate threat maps for the very first time.

Previously, we utilized AI algorithms for knowledge discovery on landslides [29, 30], global event analysis [6], person identification [31], and cardiac abnormality detection [32–34]. These studies mandated feature attributes to be existent on which the AI algorithms could execute. Therefore, for this study, feature attributes are from the news description with the help of NER. NER extracts feature attributes from news description as shown previously in [5, 6]. It should be highlighted that our previous research in [6] focused on obtaining breaking news (i.e., filtering only the most urgent news needing attention) using anomaly detection. However, in this research, all news events are analyzed (i.e., without any filters) for generating global threat map of troubled locations.

3 Material and Methods

As seen from Table 1, all the existing systems in news monitoring used only 1 or 2 news sources. For example, Twitter as a single source for monitoring new was adopted by researchers in [1–4, 12–14]. BBC News database was used by [17]. Researchers in [8] used two different news sources, namely summary of World Broadcast and New York Times. It is almost impossible to ascertain global situational awareness by only monitoring 1 or 2 news sources. Moreover, the existing research works demonstrated in [1–4, 8, 10–20] were only analyzing news reported in English language. Hence, billions of news generated in other prominent languages (e.g., Chinese, Hindi, Bengali, Arabic, etc.) were never analyzed for performing global threat analysis. Lastly, almost none of the existing literature other than [19, 20] used NER to extract location information from the analyzed news reports. Without automated recognition of the location information, it is not possible to generate threat maps.

The presented system overcomes these three major challenges (i.e., monitoring limited number of news sources, analyzing news only limited set of languages, and not extracting location information) with a number of AI-based techniques (e.g., news sensor module allowing the concurrent monitoring of thousands of news sources, language detection, and translation allowing the comprehension of news in multiple languages, and NER-based geolocation decoder for automatic extraction of location information).

The integrated system has six modules, namely news sensor, splitter, sentiment detection, news fidelity analysis, geolocation decoder, and decomposition analysis. Figure 2

shows the complete system architecture. As seen from Fig. 2, the input of the proposed system is thousands of news sources including government and non-government websites, online news agencies, and almost all the social media sites of 192 countries.

3.1 News Sensor

The news sensor senses and collects news from several thousand news sources using application programming interfaces and web scrapping technologies. Figure 3a shows the usage of Twitter API in Python language for automatically listening any hashtags containing breaking news. Whenever a tweet user posts a message containing "*Breaking News,*" the news is sensed by the Twitter API shown in Fig. 2 and passed on to the rest of the sub-system for further analysis. As mentioned earlier, our system supports a wide range of social media through the usage of their proprietary APIs that could be invoked by python (as shown in Fig. 3a). Social media messages could also be captured by listening and captured by MS Power Automate as shown in Fig. 3b [35].

In Fig. 3b, breaking news are being listened from CNN, by pointing source to "@*cnnbrk.*" As seen from Fig. 3b, whenever a breaking news is posted by CNN, the message is immediately sent to our Microsoft SQL Server database for further analyses like sentiment detection, news fidelity assessment, geolocation decoding, and decomposition analysis.

Apart from using APIs, news can also be captured using web scraping technologies by deploying powerful "M Code" within Microsoft Power Query. Figure 4 shows that our deployed "*M Code*" is sensing all the news post (breaking news or any other news). This code captures all the news from BBC Website, and *ScoreSentiment()* function is being called using Microsoft Text Analytics API [36]. The news is only retained if the news is found to be negative in nature.

3.2 Splitter

After sensing all the global news from thousands of media sources, the news messages are received by the splitter, an array consisting of both news title, *H,* and news description, *D*.

$$\text{News} = \{\{h_1, d_1\}, \{h_2, d_2\}, \{h_3, d_3\}, \dots\} \tag{1}$$

The splitter module splits the header or title information from the description; hence, two separate arrays are created consisting of headers, *H,* and Description, *D*.

$$H = \{h_1, h_2, h_3, \dots\} \tag{2}$$

$$D = \{d, d_2, d_3, \dots\} \tag{3}$$

Subsequently, these header and description entities are processed separately by sentiment analysis module.

3.3 Sentiment Analysis

The sentiment analysis module obtains the sentiment on a range of 0 to 1 for both news headers (or titles) and news descriptions. For our solution, *ScoreSentiment()* function was invoked. *ScoreSentiment()* uses Microsoft Text Analytics API [36]. Equations (4) and (5) demonstrate that sentiments for the news headers and new descriptions are obtained separately for all news from $n = 0$ to N.

$$s_{h_n} = ScoreSentiment(h_n) \tag{4}$$

$$s_{d_n} = ScoreSentiment(d_n) \tag{5}$$

ScoreSentiment() takes unstructured text as parameter and returns a decimal value between 0 and 1. Values closer to 0 indicate negative sentiment, and values near to 1 signify positive sentiment. As shown earlier in Fig. 3a, we used *ScoreSentiment()* function within *M Code*. Sentiment analysis could also be performed within a graphical user interface like MS Power Automate [35]. Figure 4 shows that sentiment analysis is being performed by our MS Power Automate-based solution. As seen from Fig. 5, if a tweet message is in English, then Sentiment is obtained directly; otherwise, the tweet text is translated and then sentiment is scored. As seen from Fig. 5b, the implementation of language detection, translation, and sentiment analysis is performed with Microsoft Cognitive Services' Text Analysis API [36].

3.4 News Fidelity

The news fidelity determines the trustworthiness of the news. In the modern era of fake news, random news from the internet cannot be trusted as the news arrives. Hence, the news fidelity module performs a polarity test on the sentiment obtained on both news title and news description. Previous studies successfully used sentiment polarity for the detection of fake news [37, 38]. Our implementation of sentiment polarity is shown in Eq. (6). With the value of news header sentiment being s_{h_n} and the value of news description sentiment being s_{d_n}, trusted news report is within a threshold value, δ.

$$TrustedNews = True, \; if \, |(s_{h_n} - s_{d_n})| < \delta \qquad (6)$$

where $0 \le s_{h_n} \le 1$ and $0 \le s_{d_n} \le 1$. During our experimentation, δ was set to 0.5 signaling a discrepancy between the content of the news and news header. The reason for this discrepancy is often news agencies are exaggerating news headers to be disproportionately negative for attracting readership [7]. However, the description of these untrusted news (i.e., $\delta > 0.5$) might be completely different.

3.5 Geolocation Decoder

The Geolocation decoder uses *EntityDetection()* that detects all the entities from a news description. The *EntityDetection()* function returns the entity type along with the detected entity. If the entity type is "*location,*" then the entity is passed to our geospatial database containing location context (i.e., province, district, country, and region) for the selected location.

$$\exists T_i^n \subseteq T | s_{l_i} > s_{l_j}, \; \forall n, \, m \subseteq \{1, 2, 3, \\ \ldots, N\} \wedge \forall i, \, j \subseteq \{1, 2, 3, \ldots\} \qquad (8)$$

The low-value AI split mode finds the most influential filter attribute condition T_i^n for which the lowest level of aggregated location sentiment (s_{l_j}) occurs, as represented by Eq. (9).

$$\exists T_i^n \subseteq T | s_{l_i} < s_{l_j}, \; \forall n, \, m \subseteq \{1, 2, 3, \\ \ldots, N\} \wedge \forall i, \, j \subseteq \{1, 2, 3, \ldots\} \qquad (9)$$

The decompression analysis enables the users to interactively perform data exploration and root-cause analysis for the impact of highly negative news on a given location. Even though decomposition tree analysis is an existing method that uses AI-based techniques for finding out root causes, any out-of-the-box AI algorithms or methods require significant level of parametrization and configuration tailored to the particular research problem as demonstrated in our recent research works [6, 29, 30, 39–41]. Similar to our works in [6,

---

Algorithm: Geolocation Decoder

Input: All the news descriptions, $D = \{d, d_2, d_3, \ldots\}$

Output: Complete location contexts containing province, district, country, region $\{l_1^{d_n}, l_2^{d_n}, l_3^{d_n}, \ldots\}$

$\{c_k, e_i^{d_n}\} = EntityDetection(d_n)$

If $c_k = 'location'$ then

$\qquad \{l_1^{d_n}, l_2^{d_n}, l_3^{d_n}, \ldots\} = GetLocationContext(e_i^{d_n})$

End If

Finally, the aggregated sentiments are applied to the detected location for generating heat maps for troubled locations. For all the news descriptions, d_n, if the description contains location $l_j^{d_n}$, then the sentiment for that description s_{d_n} is averaged for that location, as seen from Eq. (7).

$$s_{l_j} = \frac{\sum_{n=1}^N s_{d_n}}{N} \qquad (7)$$

3.6 Decomposition Analysis

Finally, the decomposition analysis module was used to perform AI-based root cause analysis. High-value AI split mode finds the most prominent filter attribute condition T_i^n for which the highest level of aggregated location sentiment (s_{l_j}) occurs, as represented by Eq. (8).

29, 30, 39–41], this research conducted multi-level parameterizations of decomposition tree analysis for determining global threats in real time which has never been reported in the literature. Figure 5 shows the deployed system, where all the modules described earlier were implemented with Microsoft technologies involving MS Office 365 Cloud, MS Power BI, MS Power Automate, MS SQL Server, and MS Azure Cloud [35, 36, 42]. The geospatial heat map generated by the negative sentiments of global news media is projected on Microsoft Bing Map. Our deployed interactive solution is publicly available within fully managed Microsoft environment in [43] (Fig. 6).
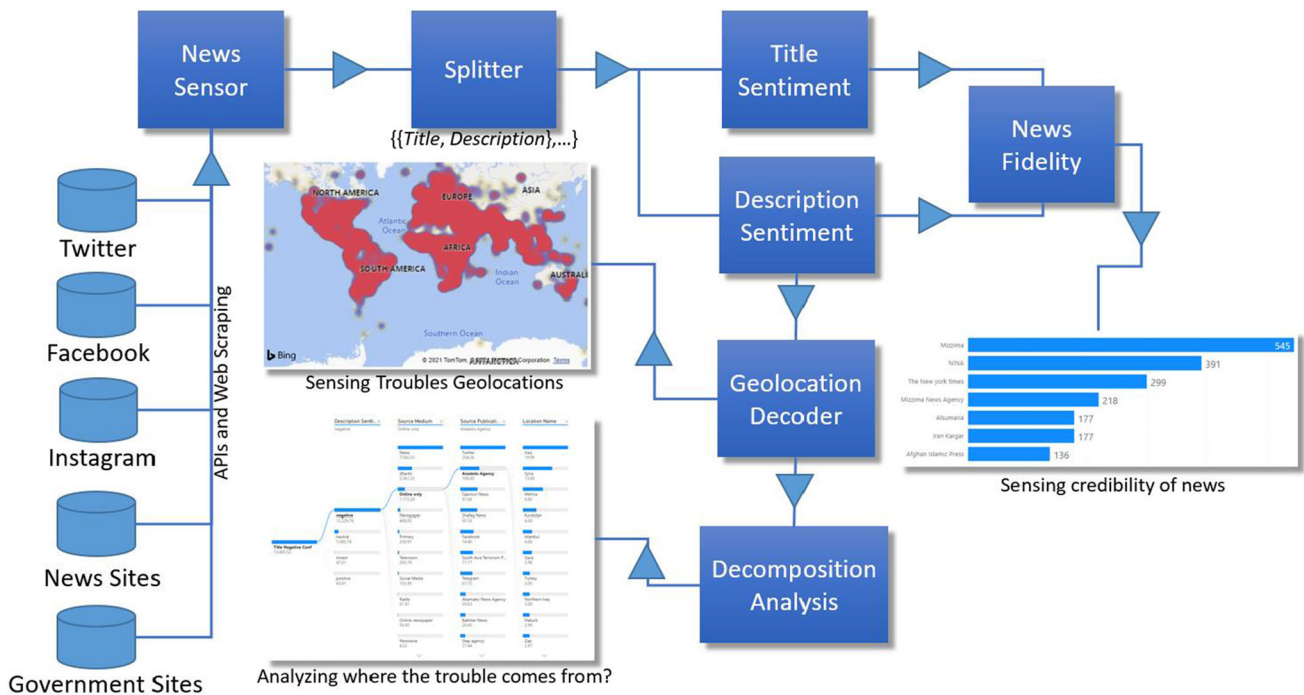
Fig. 2 System architecture for remotely sensing troubled geographic locations with global media sensors

Fig. 3 Proposed system capturing real-time social media post by the news sensor **a** Twitter API invoked in Python for listening to any hash-tags containing "Breaking News" and **b** all breaking news from CNN being captured by continuous monitoring of @cnnbrk Twitter handle with MS Power Automate

4 Result

News sensor module sensed 22,425 news reports from 2397 distinct sources between June 2, 2021, and September 1, 2021. During this short period time, global news of 192 countries were collected from government websites, news channels, Twitter, Facebook, Instagram, YouTube, etc. These 22,425 news reports were also analyzed through sentiment analysis, news fidelity analysis, geolocation decoding process, and decomposition analysis. Sentiment analysis process revealed few cases of news where the titles were negative, and the descriptions of the news were positive. As shown in Fig. 7, all titles had either positive, negative, or neutral sentiments. On the other hand, the descriptions could either have positive, negative, neutral, or mixed sentiments. Hence, as shown in Fig. 7, there could be $C(3,2)XC(4, 2)$ or 18 possible combinations of 2×2 confusion matrixes.

Therefore, Table 2 represents 18 cases of 2X2 confusion matrixes from (a) to (r). As seen from Table 2, there were 48 cases of news where the title was negative and the description was positive (i.e., (a), (c), (d), (f) from Table 2). Moreover, there were additional 48 cases where the titles of the news

BBC_WebScraping

Display Options ▼ ?

```
1   let
2       Source = Web.BrowserContents("https://www.bbc.com/"),
3       #"Extracted Table From Html" = Html.Table(Source, {{"Column1", ".module\-\-highlight .media__link"},
            {"Column2", ".module\-\-highlight .media__tag"}, {"Column3", ".module\-\-highlight
            .block-link__overlay-link"}, {"Column4", ".module\-\-highlight .media__summary"}, {"Column5",
            ".module\-\-highlight .module__title_link.tag"}}, [RowSelector=".module\-\-highlight .media__link"]),
4       #"Changed Type" = Table.TransformColumnTypes(#"Extracted Table From Html",{{"Column1", type text}, {"Column2",
            type text}, {"Column3", type text}, {"Column4", type text}, {"Column5", type text}}),
5       #"Removed Columns" = Table.RemoveColumns(#"Changed Type",{"Column5", "Column1"}),
6       #"Renamed Columns" = Table.RenameColumns(#"Removed Columns",{{"Column2", "News Category"}, {"Column3", "News
            Title"}, {"Column4", "News Description"}}),
7       #"Filtered Rows" = Table.SelectRows(#"Renamed Columns", each ([News Description] <> null)),
8       #"Applied Score sentiment" = AIFunctions.PostProcess(Table.AddColumn(#"Filtered Rows", "Score sentiment", each
            #"Score sentiment"([News Description], "en")), "Score sentiment",
            "CognitiveServices.ScoreSentimenttexttext") meta [FunctionBrowser="TextAnalytics"],
9       #"Renamed Columns1" = Table.RenameColumns(#"Applied Score sentiment",{{"Score sentiment", "Description"}}),
10      #"Applied Extract key phrases" = AIFunctions.PostProcess(Table.AddColumn(#"Renamed Columns1", "Extract key
            phrases", each #"Extract key phrases"([News Description], null)), "Extract key phrases",
            "CognitiveServices.ExtractKeyPhrasestexttext") meta [FunctionBrowser="TextAnalytics"],
11      #"Added Conditional Column" = Table.AddColumn(#"Applied Extract key phrases", "Custom", each if [Description]
            < 0.1 then "Negative News" else "Regular News")
12  in
13      #"Added Conditional Column"
```

✓ No syntax errors have been detected.

Life at 50C: Mexico's struggle for water

Climate change is causing once-mighty rivers to dry up and temperatures to rise to deadly levels in Mexico.

Extract key phrases

Negative Sentiment

change, Climate, deadly, mighty, rivers, temperatures, Mexico, levels

0.04090175

Fig. 4 Proposed system capturing real-time BBC news post by the news sensor **a** news title, and news description being captured in real time by "M Code" within MS Power Query by using web scraping technology and **b** a negative news that was captured using the M Code is displayed with news header, description, sentiment, and extracted key phrases
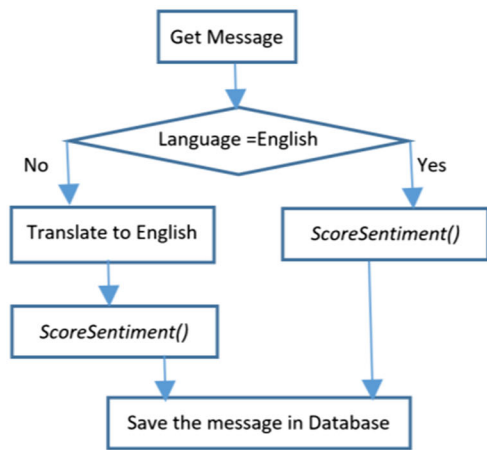
(a)
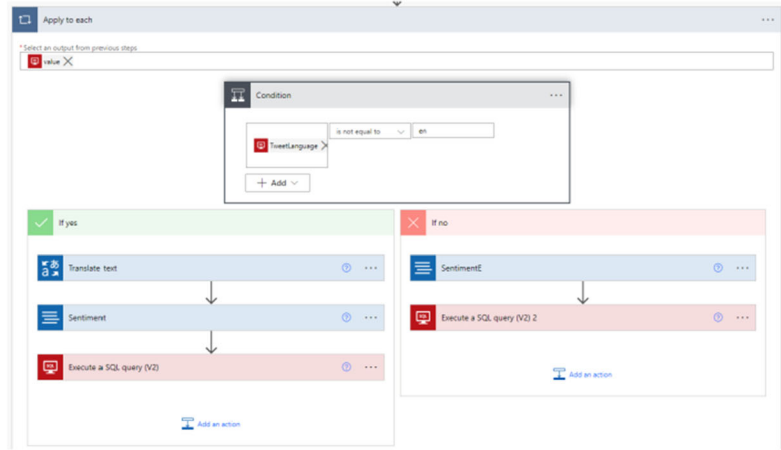
(b)

Fig. 5 The process of language detection, translation, and sentiment analysis. **a** Logical flow diagram of sentiment analysis within the proposed solution and **b** implementation of sentiment analysis in Microsoft Power Automate

were negative and the descriptions of the news were found to be of mixed sentiments (as seen from Table 2, (j), (l), (m), (o), (p), (r)). Both these cases are highlighted in Table 2 since these news reports were flagged as of low fidelity. Globally news agencies deliberately make negative headline to attract viewership as people are attracted to negative news [7–9]. However, if the headlines of the news are found to be negative and the corresponding descriptions are found to be opposite, then quite often these news reports suffer loses on credibility.

Table 3 further delves into the details of news fidelity issue with 5 different cases. Case 1 of Table 3 represents all the news flagged a low fidelity. Within case 1, there were 96 news reports from 64 media sources. Case 1 is essentially the combination of case 2 and case 3. Case 2 represents the

news reports where the titles were negative and the descriptions were positive. Within case 2, there were 48 cases of news from 34 media sources. Case 2 had the lowest level of fidelity. In case 3, all the news titles had negative sentiment and their description has mixed sentiments. Case 4 represents medium fidelity news with titles being negative and corresponding descriptions having neutral sentiments. Finally, case 5 exemplifies news reports with negative titles and negative descriptions. There were about 12,852 news reports from 1908 media sources under case 5.

Table 3 depicts threat maps for each of the cases from 1 to 5. Threat maps were generated by the deployed geolocation decoding module that extracted 11,668 location entities along with their corresponding contextual information and projecting the aggregated sentiments associated with these
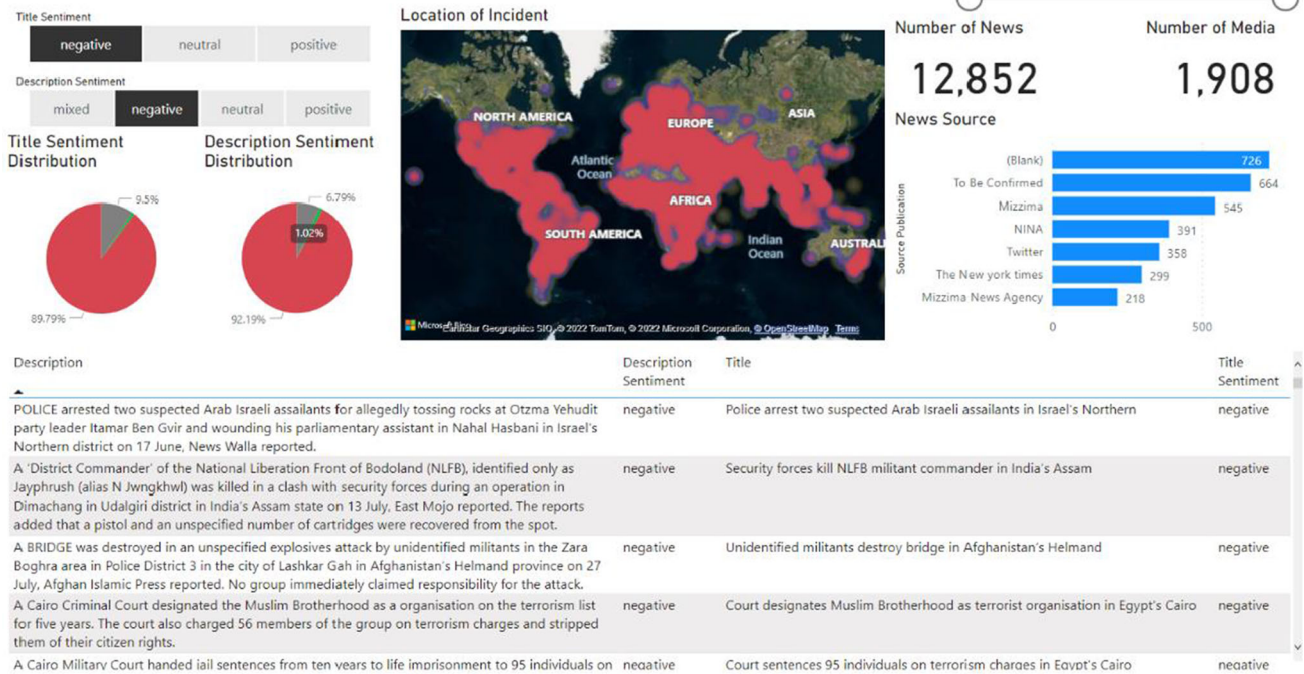
Fig. 6 System interface of the remote threat-maps generation solution (analyzing all the negative news titles with negative news descriptions with AI)
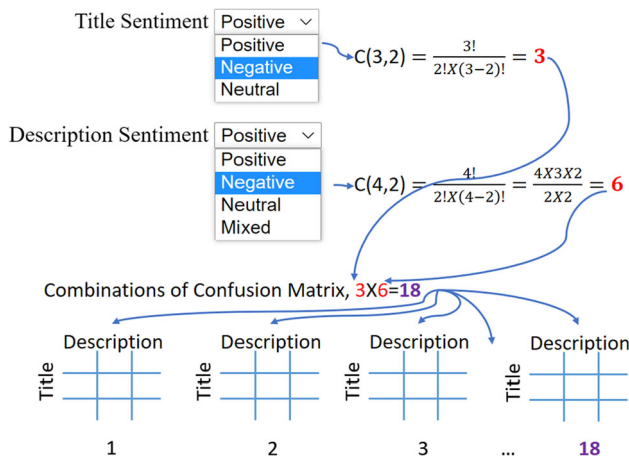

Fig. 7 Possible combinations for titles and description sentiments (i.e., 18 possible combinations)

locations on to Microsoft Bing Map. Table 3 also shows the top news media sources that were mostly responsible for propagating news for each case. For example, *Syrian Observatory for Human Rights* followed by *LM Neuquen* (from Argentina) and *Swissinfo* (from Switzerland) were the three top sources for low-fidelity news (i.e., case 1 of Table 3).

Decomposition tree visual demonstrates the ability of the user to find out the impact of source mediums, publications, and locations responsible for creating the news with highest or lowest level of negative confidence. With the help of decomposition analysis, Fig. 8 shows that *news* category of source medium generated the highest level of negative content. Further delving down in the detail reveals that *Mizzima* (Media Source) caused a higher level of negative news contents. In the next level, *Mandalay* (location) represented an extremely higher number of negative news during the monitored period. These explorations can be performed interactively on any number of scenarios and conditions. For example, Fig. 8 represents Hnutove (a village in eastern Ukraine, at 23,2 km northeast from the center of Mariupol) to be one of the most trouble-free locations within the monitored period of 92 days.

5 Discussion

This paper uses modern cloud-based architecture and APIs for performing news acquisition, sentiment analysis, entity detection, and geolocation along with a methodology on determining news fidelity and hyper-parameterization of decomposition tree analysis for solving a research problem of real-time global threat detection for the first time. Moreover, the innovative solution was evaluated extensively in terms of robustness (as shown in Eq. 10) and accuracy (as shown in Tables 2, 3, 4). Using the existing AI-based techniques and
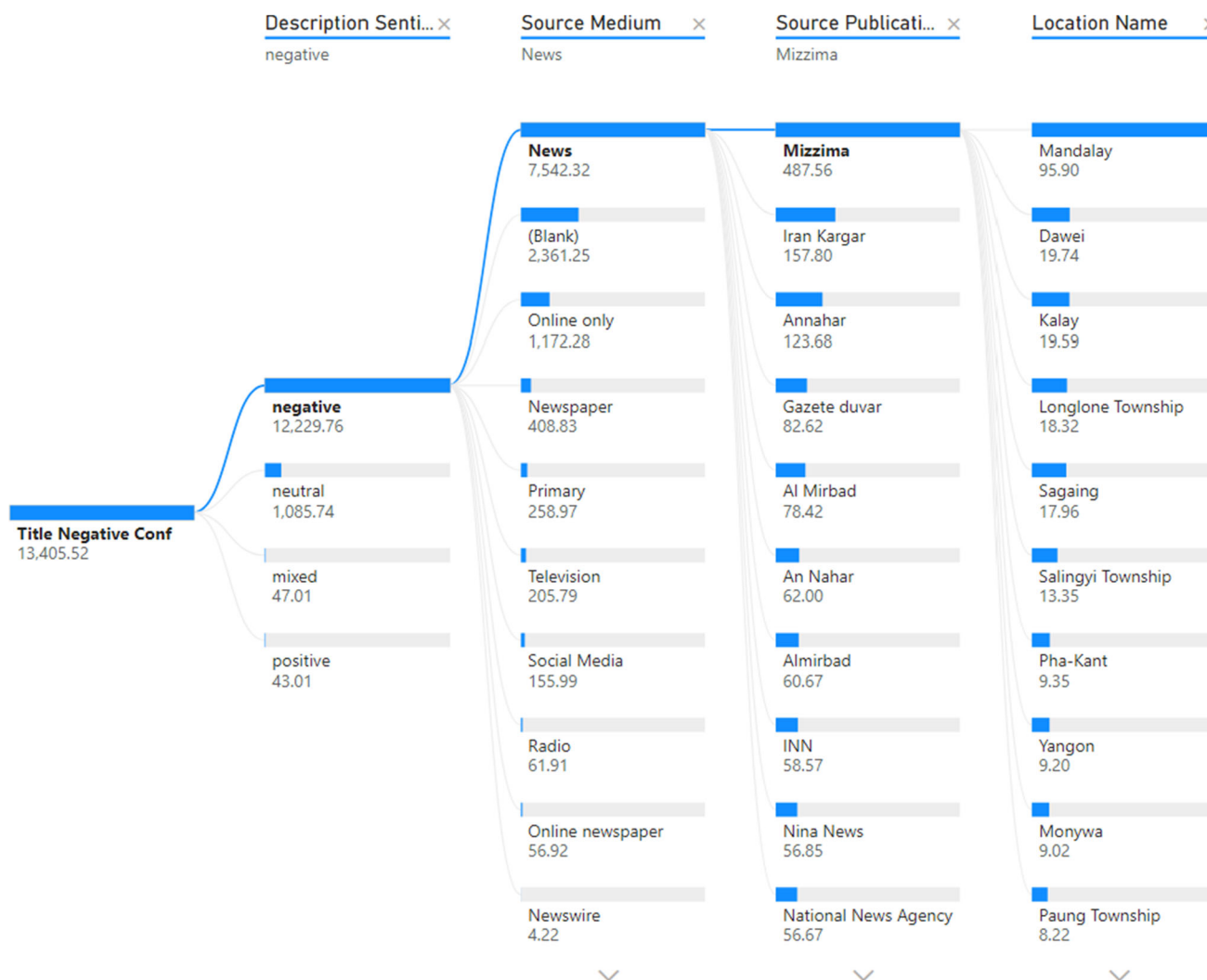
Fig. 8 Location of the top negative news

tailoring them with hyper-parameterization to solve research problems with innovative solution have been demonstrated in our research works [6, 29, 30, 39–41].

The proposed system is capable to handle a wide range of scenarios, configured by the users through the system interfaces. As seen from Fig. 6, the system provides three main inputs for configuring the scenarios: namely, option box of title sentiment, option box of description sentiment, and the date range selector. Title sentiment had three available selections (i.e., negative, neutral, and positive). Therefore, for title sentiment, there are five possible filter settings as represented by $(2^{|TitleSentiment|} - 1)$, in which the formula for calculating the power set of *Title Sentiment Attribute* minus 1 (i.e., $P(TitleSentiment) - 1$), where $|TitleSentiment|$ is the cardinality of *TitleSentiment*. 1 is deduced because the power set also includes an empty set, and the selection of an empty set is not supported by the system presented. Similarly, description sentiment had four available selections

(i.e., mixed, negative, neutral, and positive). Lastly, any dates could be selected between June 2, 2021, and September 1, 2021 (i.e., 92 days). Hence, the number or possible scenarios supported by the system could be calculated using Eq. (10).
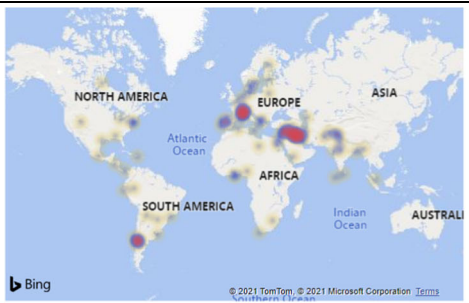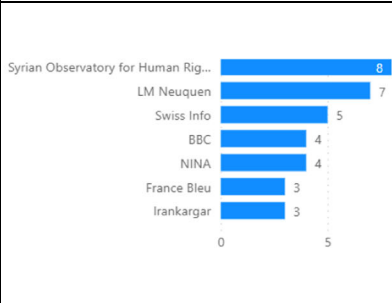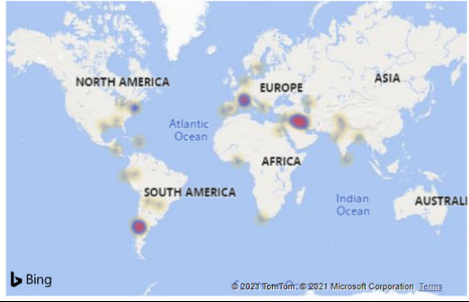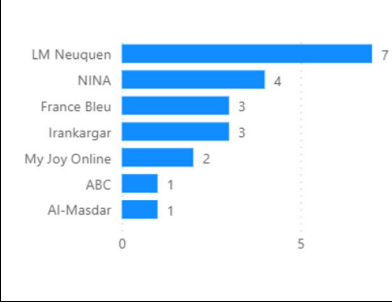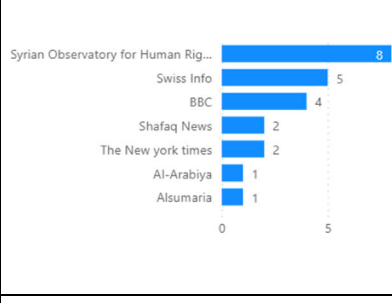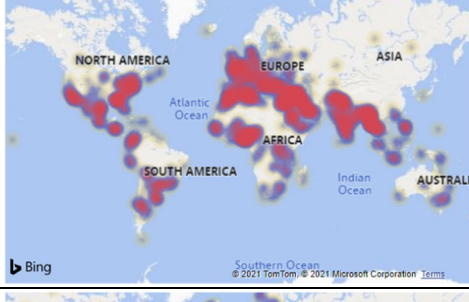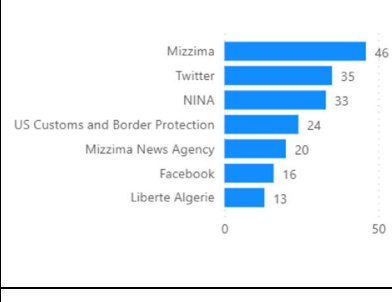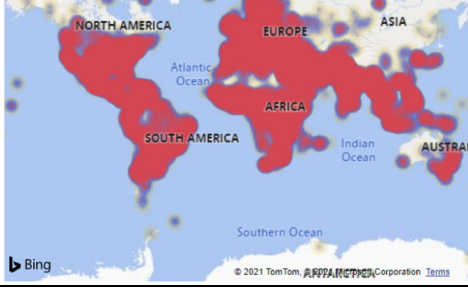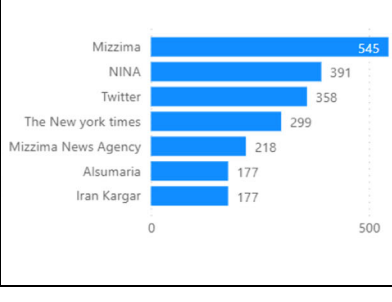
$$Scenarios = (2^{|TitleSentiment|}$$
$$- 1)X(2^{|DescriptionSentiment|} - 1)X(2^{|DateRange|}$$
$$- 1) = (2^3 - 1)X(2^4 - 1)X(2^{92} - 1) = 3.71X10^{29} \quad (10)$$

Figure 9 represents one of these scenarios that is negative titles with negative descriptions on the selected date range of June 12, 2021, to July 12, 2021. Upon this selection, the proposed system filtered 4443 news from 958 news sources, representing the selected scenario and generated the trouble threat map in Microsoft Bing Map for the chosen scenario (as seen from Fig. 9).

Table 2 Eighteen confusion matrix combinations from (a) to (r)

(a)

	Description	
	Positive	Negative
Positive	10	72
Negative	**48**	12852

(b)

	Description	
	Positive	Negative
Positive	10	72
Neutral	37	5198

(c)

	Description	
	Positive	Negative
Neutral	37	5198
Negative	**48**	12852

(d)

	Description	
	Positive	Neutral
Positive	10	50
Negative	**48**	1045

(e)

	Description	
	Positive	Neutral
Positive	10	50
Neutral	37	3042

(f)

	Description	
	Positive	Neutral
Neutral	37	3042
Negative	**48**	1045

(g)

	Description	
	Neutral	Negative
Positive	50	72
Negative	1045	12852

(h)

	Description	
	Neutral	Negative
Positive	50	72
Neutral	3042	5198

(i)

	Description	
	Neutral	Negative
Neutral	3042	5198
Negative	1045	12852

(j)

	Description	
	Negative	Mixed
Positive	72	3
Negative	12852	**48**

(k)

	Description	
	Negative	Mixed
Positive	72	3
Neutral	5198	20

(l)

	Description	
	Negative	Mixed
Neutral	5198	20
Negative	12852	**48**

(m)

	Description	
	Neutral	Mixed
Positive	50	3
Negative	1045	**48**

(n)

	Description	
	Neutral	Mixed
Positive	50	3
Neutral	3042	20

(o)

	Description	
	Neutral	Mixed
Neutral	3042	20
Negative	1045	**48**

(p)

	Description	
	Positive	Mixed
Positive	10	3
Negative	48	**48**

(q)

	Description	
	Positive	Mixed
Positive	10	3
Neutral	37	20

(r)

	Description	
	Positive	Mixed
Neutral	37	20
Negative	**48**	**48**

Table 3 Threat maps of locations with sources of media for news of different fidelity

	Description	News reports effected	News source effected	Threat map showing troubled locations for each case	Top sources of reported news for each case
Case 1 (Low Fidelity)	All news sensed with sentiment polarity (i.e., case 1 and case 2 combined)	96	64		
Case 2 (Lowest Fidelity)	News sensed with Negative titles and Positive descriptions	48	34		
Case 3 (Low Fidelity)	News sensed with Negative titles and mixed descriptions	48	30		
Case 4 (Medium Fidelity)	News sensed with Negative titles and neutral descriptions	1045	388		
Case 5 (High Fidelity)	News sensed with negative titles and negative descriptions	12852	1908		
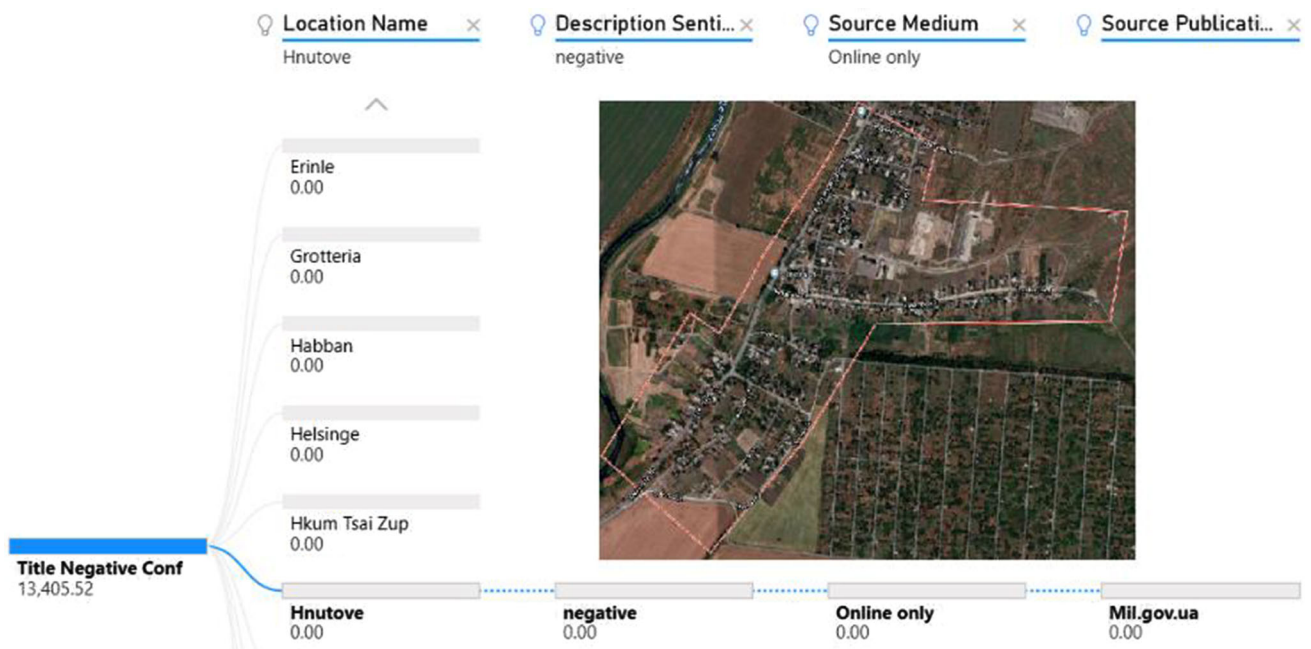
Fig. 9 Most trouble-free location sensed by the deployed system from media source mil.gov.ua

All the modules used within the presented system were comprehensively evaluated in terms of true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), accuracy, precision, recall, and F1-score. The detailed results with all the evaluation metrics used are presented in Table 4. Accuracy provides the overall accuracy of the model, and it is represented by the fraction of the total samples that were correctly classified the by the classifier, as seen in Eq. (11).

Precision informs what fraction of predictions as a positive class was positive (shown in Eq. (12)).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{12}$$

Recall informs what fraction of all positive samples was correctly predicted as positive by the classifier as shown in Eq. (13).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{11}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{13}$$

Table 4 Performance evaluations of the classification algorithms

	TP	TN	FP	FN	Precision	Recall	F1-Score	Accuracy
News fidelity—low fidelity	96	4	3	5	0.9697	0.95	0.96	0.92593
News fidelity—high fidelity	12,852	117	317	34	0.97593	0.997	0.9865	0.97365
News fidelity—moderate fidelity	9573	35	23	12	0.9976	0.999	0.9982	0.99637
Title sentiment—negative	13,993	234	23	54	0.99836	0.996	0.9973	0.99462
Title sentiment—positive	135	3	8	3	0.94406	0.978	0.9609	0.92617
Title sentiment—neutral	8297	29	43	12	0.99484	0.999	0.9967	0.99344
Description sentiment—negative	18,122	642	124	310	0.9932	0.983	0.9882	0.97739
Description sentiment—positive	95	12	10	6	0.90476	0.941	0.9223	0.86992
Description sentiment—neutral	4137	43	28	53	0.99328	0.987	0.9903	0.98099
Description sentiment—mixed	71	12	2	8	0.9726	0.899	0.9342	0.89247
Geolocation decoder	11,668	120	56	41	0.99522	0.996	0.9959	0.99184
Total	79,039	1251	637	538	0.99201	0.993	0.9926	0.98558
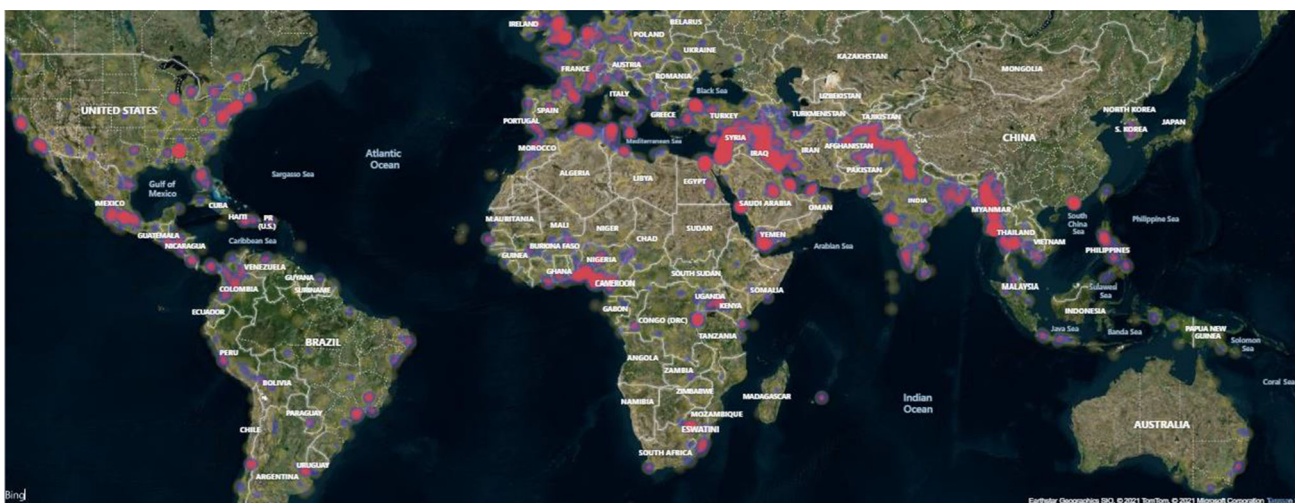
Fig. 10 Negative titles with negative descriptions from June 12, 2021, to July 12, 2021 (4443 news from 958 news sources)
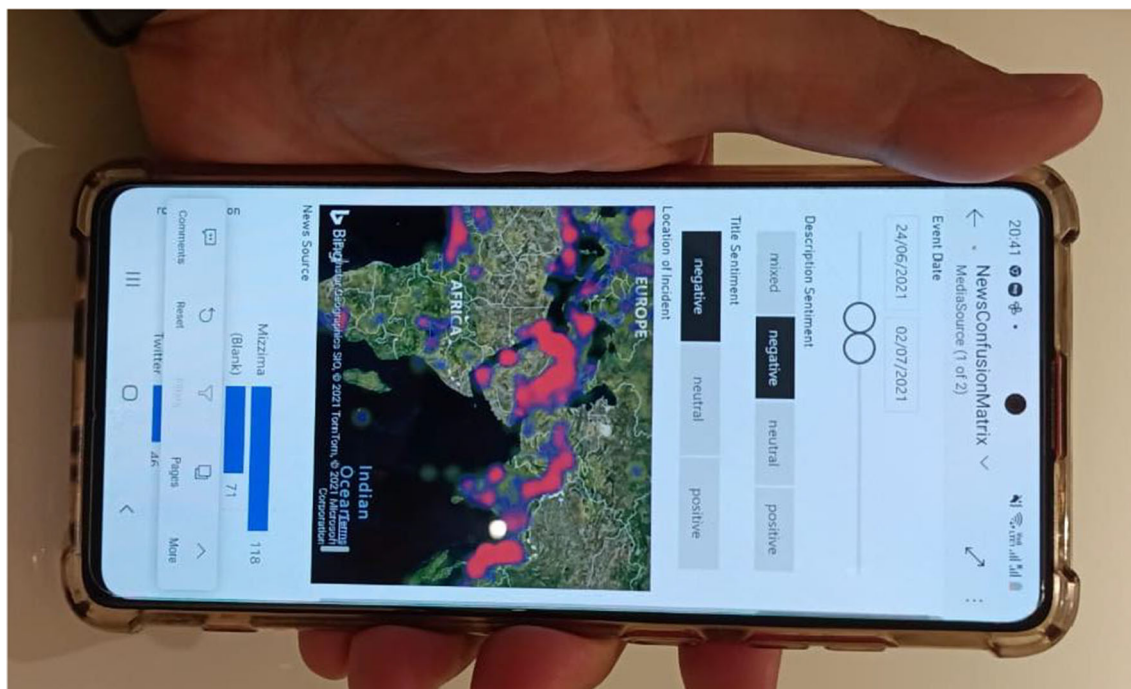
Fig. 11 A user remotely viewing troubled locations on a Samsung Note 10 mobile device from June 24, 2021, to July 02, 2021, where the news titles and news descriptions are both negative

F1-score combines precision and recall into a single measure as shown in Eq. (14).

$$F_1 Score = \frac{2 X \, Precision \, X \, Recall}{Precision + Recall} = \frac{2 X TP}{2 X TP + FP + FN} \quad (14)$$

As seen from Table 4, the overall classification accuracy of the entire system was evaluated as 0.992 for precision, 0.993 for recall, 0.993 for F1-score, and 0.985 for accuracy. The breakdown of all the individual class accuracy is also depicted in Table 4. Our implementation of the geolocation

decoder had the highest accuracy (0.992 accuracy, 0.996 F1-score, 0.996 recall, and 0.995 precision).

Finally, the entire system was deployed in mobile environment (i.e., both in Android and iOS platforms) for its capacity to serve an intelligence office or strategic decision maker who might be located at a remote location (as seen from Fig. 10). Since the chosen platforms (including Microsoft Power BI, Microsoft Power Automate, and Microsoft Azure) operate seamlessly in mobile environment, the presented system was successfully deployed in mobile and tablet environment on both Android and iOS ecosystem. Figure 11

shows the proposed system deployed in Samsung Note 10 and the user looking at a threat map with the following conditions: (1) Dates between June 24, 2021, and July 02, 2021; (2) negative news titles and (3) negative news descriptions. Being a cloud-based solution, the computational overhead of the presented solution is shared across the Microsoft Cloud (Microsoft Power Platform and Office 365), and the client system (i.e., Mobile or desktop apps running in iOS, Android, and Windows). The scalable nature of cloud allows complex computational load to be seamlessly distributed across multiple resources as required. The complete architecture of such system involving Microsoft Power BI, Microsoft Power Automate (Fig. 1), and Microsoft SQL Server is demonstrated in our recent works in [6, 44].

The presented system could interactively generate threat maps from a wide range of $3.71 X 10^{29}$ scenarios on several platforms covering desktop, cloud, and mobile environment with an overall accuracy of 0.985. The interactive system is publicly accessible from fully managed Microsoft environment in [43]. The complete source code for the developed system is located in [45].

Experiment was performed with very large set of geographically dispersed (192) countries and news sources (2397) in many languages. Moreover, the study was performed for a relatively longer duration (i.e., 3 months) compared to many studies reported in the literature (e.g., researchers in [46] monitored Twitter for 87 days from December 13, 2019, to March 9, 2020). This study could easily be applied to a substantially larger dataset of hundreds of thousands, and scalable natural API could easily adjust to the higher level of processing requirements for generating millions of threat maps across the globe. Hence, the presented study demonstrates strong external validity.

When new news sources understand that they might be monitored, they could change their tone of reporting the news. Changing the tone of news would directly impact the sentiment analysis process, effecting the presented news monitoring and aggregation system. Therefore, "reactive effects of the monitored source" could pose a threat to external validity.

The web scraping methods using M Codes and Microsoft Power Query are dependent under the assumption that the format of the monitored web sources (i.e., the 2397 monitored sources) remains the same. Hence, major changes in website features and updates to the formats could pose threat to internal validity during experimentation (i.e., 3 month of monitoring period).

The core features and innovations demonstrated in this paper in comparison with the existing literatures are highlighted in Table 5.

Table 5 Core contributions of the presented studies in comparison with the existing literature

Core features	Proposed approach	Comparison in existing literature
Number of languages supported for sentiment analysis	Support more than 110 languages with Microsoft Cognitive Services [47]	Only English was supported in [1– 4, 8, 10–20]
Location extractions with NER	As shown in Table 4, the accuracy of NER based geolocation decoder was found to be 0.99522, 0.996, 0.996, and 0.992 in terms of precision, recall, f1-score, and accuracy	F-score of only 0.57 demonstrated in [19]. On the other hand, 0.85 precision and 0.72 recall as evaluated in [20]
Number of news sources monitored	Concurrent monitoring of 2397 news sources	1 or 2 news sources monitored as shown in [1– 4, 8, 10–20]
Root-cause analysis with decomposition analysis	Fully supported	Not supported in [1– 4, 8, 10–20]
Mobile app support	All major platforms supported (e.g., iOS, Android, Windows)	Not supported in [1– 4, 8, 10–20]

6 Conclusion

In our earlier study [6], we utilized NER to extract locations from global news reports and then used anomaly detection to filter out only the most urgent news requiring attention of the decision maker. In that study [6], we envisioned creating an automated system for generating the threat maps representing troubled region. Hence, this paper is an extension of [6], where we extracted location information from all the global news to create heat map or threat map of troubled location. Fusing location information provides contextual information to the tweet message, and as a result, the accuracy and performance of sentiment analysis process increased drastically, as shown in [48].

In summary, this paper presents an innovative remote sensing application for generating global threat map of from robust scenario space with very high accuracy. The solution relied on API and web scraping-based *news sensor* module that automatically listens to thousands of global media sources for troubled locations. The presented architecture makes the proposed system execute in complete autonomous mode without the need for manual data integration, cleansing, or modelling. A geospatial intelligence officer or strategic decision maker can quickly create a scenario using the deployed mobile app or desktop computer, and immediately, the right threat maps are generated and presented to the user. Therefore, the user of the system can remotely view the global troubled locations from any device of choice (iOS, Android, Windows, etc.).

The system was deployed in Microsoft Cloud environment and tested from June 2, 2021, to September 1, 2021. During this time, the proposed system autonomously sensed global events from 2,397 media sources (e.g., government websites and news agencies of 192 countries as well as all major social media sites that include Facebook, Instagram, Twitter, etc.). About 22,425 news events were automatically analyzed with AI-based services. The system could generate threat maps on any number of scenarios out of $3.71 X 10^{29}$ possible scenario spaces that included automatically classified and verified 11,668 worldwide locations. The evaluation of classification for the entire system was found to be of 0.992 precision, 0.993 recall, 0.993 F1-score, and 0.985 accuracy.

According to the literature and to the best of our knowledge, this is the first reported study that generated global threat maps in complete autonomous mode without the assistance of a data scientist. The end-to-end automation makes the system highly suitable for a strategic decision maker who wants to remain updated about any trouble happening anywhere in the world.

In this study, we did not classify the news in terms of political, sports, domestic, international, terrorism, and any other criteria. In future, we endeavor to autonomously classify the news types, hence allowing the users to obtain threat maps belonging to specific news criteria (e.g., threat map related to terrorism activities). Moreover, in future, we would like to incorporate encryption for making the proposed system secured against incoming cyber threats [49, 50] and try involving new mobile-based AI technologies [51]. These AI-based automated threat maps can be very useful for Hajj applications affecting people privacy from different locations gathered in Makkah [52], benefitting from IoT [53], and serving the international community [54].

References

1. Park, S.B.; Kim, J.; Lee, Y.K.; Ok, C.M.: Visualizing theme park visitors' emotions using social media analytics and geospatial analytics. Tour. Manage. **80**, 104127 (2020)
2. Qi, W.; Procter, R.; Zhang, J.; Guo, W.: Mapping consumer sentiment toward wireless services using geospatial twitter data. IEEE Access **7**, 113726–113739 (2019)
3. Middleton, S.E.; Middleton, L.; Modafferi, S.: Real-time crisis mapping of natural disasters using social media. IEEE Intell. Syst. **29**, 9–17 (2014)
4. Vashisht, G.; Sinha, Y.N.: Sentimental study of CAA by location-based tweets. Int. J. Inf. Technol. **13**, 1555–1567 (2021)
5. Hoang, T.B.N.; Mothe, J.: Location extraction from tweets. Inf. Process. Manage. **54**(2), 129–144 (2018)
6. Sufi, F.K.; Alsulami, M.: Automated multidimensional analysis of global events with entity detection, sentiment analysis and anomaly detection. IEEE Access **9**, 152449–152460 (2021)
7. Trussler, M.; Soroka, S.: Consumer demand for cynical and negative news frames. The Int. J. Press/Polit. **19**(3), 360–379 (2014)
8. Leetaru, K. H.: Culturomics 2.0: Forecasting large-scale human behavious using global news media tone in time and space. First Monday. **16**(9) (2011)
9. Pinker, S.: The media exaggerates negative news. This distortion has consequences (2018)
10. Islam, M.; Kabir, M.; Ahmed, A.; Kamal, A.; Wang, H.; Ulhaq, A.: Depression detection from social network data using machine learning techniques. Health Inf. Sci. Syst. **6**(1) (2018)
11. Ricard, B. J.; Marsch, L. A.; Crosier, B.; Hassanpour, S.: Exploring the utility of community-generated social media content for detecting depression: an analytical study on Instagram. J. Med. Internet Res. **20**(12) (2018)

12. Al-twairesh, N.; AL-negheimish, H.: Surface and deep features ensemble for sentiment analysis of arabic tweets. IEEE Access, pp. 84122–84131 (2019)

13. Park, C. W.; Seo, D. R.: Sentiment analysis of Twitter corpus related to artificial intelligence assistants. In: 5th international conference on industrial engineering and applications (ICIEA) (2018)

14. Ebrahimi, M.; Yazdavar, A. H. ; Sheth, A.: Challenges of sentiment analysis for dynamic events. IEEE Intell. Syst. **32**(5) (2017)

15. Vanaja, S.; Belwal, M.: Aspect-level sentiment analysis on E-commerce data. In: International Conference on Inventive Research in Computing Applications (ICIRCA 2018) (2018)

16. Zvarevashe, K.; Olugbara, O. O.: A framework for sentiment analysis with opinion mining of hotel reviews. In: Conference on Information Communications Technology and Society (ICTAS) 2018 (2018)

17. Shirsat, V. S.; Jagdale, R. S.; Deshmukh, S. N.: Document level sentiment analysis from news articles. In: International Conference on Computing, Communication, Control and Automation (ICCUBEA) (2017)

18. Li, J.; Qiu, L.: A sentiment analysis method of short texts in microblog. In: IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC) (2017)

19. Chen, X.; Faviez, C.; Schuck, S.; Lillo-Le-Louët, A.; Texier, N.; Dahamna, B.; Huot, C.; Foulquié, P.; Pereira, S.; Leroux, V.; Karapetiantz, P.; Guenegou-Arnoux, A.; Katsahian, S.; Bousquet, C.; Burgun, A.: Mining patients' narratives in social media for pharmacovigilance: adverse effects and misuse of methylphenidate. Front. Pharmacol. **9**(541) (2018)

20. Cameron, D.; Smith, G. A.; Daniulaityte, R.; Sheth, A. P.; Dave, D.; Chen, L.; Anand, G.; Carlson, R.; Watkins, K. Z.; Falck, R.: PREDOSE: a semantic web platform for drug abuse epidemiology using social media. J. Biomed. Inf. **46**(6) (2013)

21. Pang, B.; Lee, L.; Vaithyanathan, S.: Thumbs up?: Sentiment classification using machine learning techniques. In: Conf. Empirical Methods Natural Lang. Process. (2002)

22. Turney, P. D.: Thumbs up or thumbs down?: Semantic orientation applied. In: 40th Annu. Meeting (2002)

23. Naseem, U.; Razzak, I.; Khushi, M.; Eklund, P. W.; Kim, J.: COVIDSenti: a large-scale benchmark Twitter. IEEE Trans. Comput. Soc. Syst. (2020)

24. Li, L.; Zhang, Q.; Wang, X.; Zhang, J.: Characterizing the propagation of situational information in social media during COVID-19 epidemic: a case study on weibo. IEEE Trans. Comput. Soc. Syst. **7**(2), 556–562 (2020)

25. Chan, B.; Lopez, A.; Sarkar, U.: The canary in the coal mine tweets: social media reveals public perceptions of non-medical use of opioids. PLOS One (2015)

26. McNaughton, E.C.; Black, R.A.; Zulueta, M.G.; Budman, S.H.; Butler, S.F.: Measuring online endorsement of prescription opioids abuse: an integrative methodology. Pharmacoepidemiol. Drug Saf. **21**(10), 1081–1092 (2012)

27. Mäntylä, M.V.; Graziotin, D.; Kuutila, M.: The evolution of sentiment analysis - a review of research topics, venues, and top cited papers. Comput. Sci. Rev. **27**, 16–32 (2018)

28. E. Batbaatar and K. H. Ryu, "Ontology-Based Healthcare Named Entity Recognition from Twitter Messages Using a Recurrent Neural Network Approach," *International Journal of Environmental Research and Public Health,* vol. 16, no. 3628, 2019.

29. Sufi, F. K.; Alsulami, M.: Knowledge discovery of global landslides using automated machine learning algorithms. IEEE Access, **9** (2021)

30. Sufi, F. K.: AI-Landslide: software for acquiring hidden insights from global landslide data using artificial intelligence. Softw. Impacts, **10**, 100177 (2021)

31. Sufi, F.; Khalil, I.: Faster person identification using compressed ECG in time critical wireless telecardiology applications. J. Netw. Comput. Appl. **34**(1), 282–293 (2011)

32. Sufi, F.; Fang, Q.; Khalil, I.; Mahmoud, S.S.: Novel methods of faster cardiovascular diagnosis in wireless telecardiology. IEEE J. Sel. Areas Commun. **27**(4), 537–552 (2009)

33. Sufi, F.; Khalil, I.: Diagnosis of cardiovascular abnormalities from compressed ECG: a data mining-based approach. IEEE Trans. Inf Technol. Biomed. **15**(1), 33–39 (2010)

34. Sufi, F.; Khalil, I.: A clustering based system for instant detection of cardiac abnormalities from compressed ECG. Expert Syst. Appl. **38**(5), 4705–4713 (2011)

35. Microsoft Documentation: Microsoft power automate (2021). [Online]. https://docs.microsoft.com/en-us/power-automate/. (Accessed 29 August 2021)

36. Microsoft Documentation: Text analytics api documentation (2021). [Online]. https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/. (Accessed 3 Aug 2021)

37. Samonte, M. J. C.: Polarity analysis of editorial articles towards fake news detection. In: ICIEB '18: Proceedings of the 2018 International Conference on Internet and e-Business (2018)

38. Ajao, O.; Bhowmik, D.; Zargari, S.: Sentiment aware fake news detection on online social networks. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2019)

39. Sufi, F.: AI-GlobalEvents: a software for analyzing, identifying and explaining global events with artificial intelligence. Softw. Impacts **11**(100218), 1–5 (2022)

40. Sufi, F.; Khalil, I.: Automated disaster monitoring from social media posts using AI based Location Intelligence and sentiment analysis. IEEE Trans. Comput. Soc. Syst. https://doi.org/10.1109/TCSS.2022.3157142) (2022)

41. Sufi, F.; Razzak, I.; Khalil, I.: Tracking anti-vax social movement using ai based social media monitoring. IEEE Trans. Technol. Soc. **4**. https://doi.org/10.1109/TTS.2022.3192757 (2022)

42. Microsoft Documentation: Model interpretability in Azure Machine Learning (2021). [Online]. https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability. (Accessed 15 08 2021).

43. Sufi, F.: Remotely sensing troubled geographic locations with global media sensors. [Online]. https://app.powerbi.com/view?r=eyJrIjoiODdlYWU0M2ItOTQ3NS00OWJjLWE3MGQtNTI2NmFkZTEyYTg0IiwidCI6IjBkMWI4YmRlLWZmYzEtNGY1Yy05NjAwLTJhNzUzZGFjYmEwNSJ9. (Accessed 29 September 2021).

44. Sufi, F.K.: Identifying the drivers of negative news with sentiment, entity and regression analysis. Int. J. Inf. Manage. Data Insights **2**(1), 100074 (2022)

45. Sufi, F.: Analysis of global events solution source files. (2021) [Online]. https://github.com/DrSufi/GlobalEvent. (Accessed 27 Oct 2021).

46. Boon-Itt, S.; Skunkan, Y.: Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. JMIR Public Health Surveill. **6**(4), e21978 (2020)

47. Microsoft Documentation: Languages supported by Language Detection. 2 Nov 2021. [Online]. https://docs.microsoft.com/en-us/azure/cognitive-services/language-service/language-detection/language-support. (Accessed 6 May 2022)

48. Lim, W. L.; Ho, C. C.; Ting, C.-Y.: Sentiment analysis by fusing text and location features of geo-tagged tweets. IEEE Access **8** (2020)

49. Altalhi, S.; Gutub, A.: A survey on predictions of cyber-attacks utilizing real-time twitter tracing recognition. J. Ambient. Intell. Humaniz. Comput. **12**, 10209–10221 (2021)

50. Alkhudaydi, M.; Gutub, A.: Securing data via cryptography and arabic text steganography. SN Comput. Sci. **2**(46) (2021)

51. Singh, A.; Satapathy, S. C.; Roy, A.; Gutub, A.: AI based mobile edge computing for IoT: applications, challenges, and future scope. Arab. J. Sci. Eng. (2022). https://doi.org/10.1007/s13369-021-06348-2

52. Shambour, M. K.; Gutub, A.: Personal privacy evaluation of smart devices applications serving Hajj and Umrah Rituals. J. Eng. Res. (2021). https://doi.org/10.36909/jer.13199

53. M. K. Shambour, A. Gutub (2022) Progress of IoT research technologies and applications serving Hajj and Umrah. Arab. J. Sci. Eng.; 47: 1253–1273. https://doi.org/10.36909/jer.13199

54. Farooqi, N.; Gutub, A.; Khozium, M.O.: Smart community challenges: enabling IoT/M2M technology case study. Life Sci J **16**(7), 11–17 (2019)