Harvesting social media with artificial intelligence for medical threats mapping and analytics

Walter David [a,*], Michelle King-Okoye [a], Alessandro Capone [b], Gianluca Sensidoni [b], Silvia Elena Piovan[c]

[a] *Ronin Institute, walter.david@ronininstitute.org, michelle.king-okoye@ronininstitute.org*
[b] *Expert.ai, acapone@expert.ai, gsensidoni@expert.ai*
[c] *University of Padua, silvia.piovan@unipd.it*

* Corresponding author

Abstract: The COVID-19 pandemic has exposed both national and organizational vulnerabilities to infectious diseases and has impacted, with devastating effects, many business sectors. Authors have identified an urgent need to effectively plan for future threats, by exploiting emerging technologies to forecast, predict and anticipate action at the strategic, operational and local level thus strengthening the capacity of national and international responders. In order to do this, we need an approach to increase awareness of actors involved. The purpose of this study is to investigate how improved medical intelligence, harvesting from big data available from social media, scientific literature and other resources such as local press, can improve situational awareness to take more informed decision in the context of safeguarding and protecting populations from medical threats. This paper focuses on the exploitation of large unstructured data available from microblogging service Twitter for mapping and analytics of health and sentiment situation. Authors tested an explainable artificial intelligence (AI) supported medical intelligence tool on a scenario of a megacity by processing and visualizing tweets on a GIS map. Results indicate that explainable AI provides a promising solution for measuring and tracking the evolution of disease to provide health, sentiment and emotion situational awareness.

Keywords: social media, artificial intelligence, medical intelligence, health and sentiment mapping

1. Introduction

The COVID-19 pandemic has exposed the fragility and vulnerability of health infrastructures globally and the challenges scientific advisors and public health experts face to understand the magnitude and spread of this highly infectious disease. This is well demonstrated in the time taken to develop COVID-19 vaccines, by which millions of lives succumbed to this dreadful disease. While we are still within a pandemic stage and slowly recovering as we are optimistic that vaccinations will surely bring about some normalcy, it is important to focus on prevention planning. Our goals should focus on speed, efficiency, cartography and sustainability. This could be accomplished by combining medical intelligence, artificial intelligence, deep semantics and deep learning techniques, social media, communications technologies and visualisation mapping. We have identified an urgent need to effectively prepare for countering future threats and safeguard people, by harvesting emerging technologies to forecast, predict and anticipate action at the strategic, operational and local level thus strengthening the response capacity. To do this, we need a novel approach to increase the awareness of decision-makers.

In healthcare, the management of large volumes of health data is challenging because time is crucial for diagnosis and treatment (Dash et al., 2019).

Managing large volumes of health data is regarded as the most challenging for the healthcare industry, as timely diagnosis and treatment is critical to analysis and management of data for service users, practitioners and all stakeholders involved (Dash et al., 2019). Big data (structured and unstructured) poses few challenges, apparent in most areas of medicine (Stylianou et al, 2017). Moreover, it is estimated that roughly 80% of clinical data is in an unstructured form (Linguamatics, 2021a). The human brain has difficulty to process and keep up with large amounts of information, meaning that often timely diagnosis and treatment are delayed. The analysis of big digital data can contribute to improve disease surveillance and epidemic intelligence (O'Shea, 2017). It is therefore important to make the most of efficient ways to capture and manage big data analytics for today and for the next future. Authors are involved in initiatives under the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) umbrella, to investigate how improved medical intelligence, harvesting from large unstructured data available from social media, in addition to resources available to analysts from scientific literature, national and local press, can improve situational awareness. This could prove useful also to support planning and the conduct of humanitarian operations when medical threats add complexity to the mission. This paper shows how explainable artificial intelligence (AI) can be embedded in a medical intelligence tool to improve measuring and tracking the evolution of diseases for updated health, sentiment and emotion situational awareness. As a proof of concept, the authors' approach combines Twitter microblogging service with the medical intelligence platform (MIP) developed by Expert,ai enterprise for

visual analytics, geospatial social media analytics to identify, alert and support decision making.

A scenario of a megacity, London, has been built and geotagged tweets generated from a specific geographic location have been processed by application of an artificial intelligence (AI) supported medical intelligence platform. In this paper, authors present a novel approach towards medical intelligence (MEDINT) to harvest from user-generated data, available through Twitter (Prieto, 2014). Results indicate that such approach, exploiting real-time data sharing, provides a promising solution for measuring and tracking the evolution of disease in society, early detection of disease outbreaks and support decision-making for appropriate response to medical threats.

2. Social media and medical intelligence

Social media are a promising source of data generated from local people who experience real-world phenomena and can play a crucial role during a crisis. Authors selected Twitter for the consolidated role it has played in many crises where it has proven to be an important platform for providing services and information (Kusumasari et al., 2020) and reach people in need. For example, Huang et al. (2020, 2021) demonstrate the potentials in using geotagged tweets, among other sources of data, to examine human mobility through the pandemic.

2.1 Twitter

The most widespread and worldwide used microblogging social media, Twitter's significant role in medical intelligence has come to light during the COVID-19 pandemic. This is due to the quantity of data shared every day by millions, and the impact of tweets from institutions and influencers on their followers (Murthy, 2018). Twitter can be utilised for safeguarding against medical threats due to its speed and possibility to disseminate information in emergencies and disasters preparation (Seddighi et al., 2020), to send warnings and identify needs and required responses. Its reach makes it also an ideal platform for planning and preparedness coordination (Kusumasari, et al, 2020). In fact, Twitter has been largely used as a communication tool, during flood events (Eilandera et al., 2016) when affected people tweet about their situation, what they need and how deep is the flooding water. Tweets can contain a reference to a location, and even a photo (Eilandera et al., 2016), Multiple observations from citizens reporting the same flood severity increase the probability of the observations being correct; these observations can be used to support decision (Eilandera et al., 2016). Aristotle is credited as the first person to write about the *wisdom of the crowd* in his work titled *Politics* "*it is possible that the many, though not individually good men, yet when they come together may be better, not individually but collectively, than those who are so, just as public dinners to which many contribute are better than those supplied at one man's cost"*. Flood maps were produced for a pilot project in Jakarta, a city suffering from frequent recurring floods (Eilandera et al., 2016) but social networks and blogs are increasingly being used also to disseminate personal health information. Twitter highlights public perceptions of disease outbreaks.

2.2 Medical Intelligence

The term *medical intelligence* (MEDINT) describes a functional area of intelligence, encompassing an evaluation of factors that pose any significant threat to health services (US Department of the Army, 1989). It also ensures safeguarding against medical threats (Kaufman, 2001; Bowsher et al., 2016) as, for example, the COVID-19 pandemic. MEDINT assesses threat to the global community and securitisation of diseases (Curley and Herrington, 2011) diseases that were referred to as a 'national security threat' (White House Briefing, 2014). In recent times, MEDINT has helped, with rapid assessment, in controlling and containing Ebola, SARS, H1N5 and Zika (Reperent and Osterhaus, 2019).

2.3 AI and disease surveillance

In crisis management, it is crucial to have an avant-garde information-seeking capability revealing where data is transformed into information. Faster detection of disease outbreaks is critical to surveillance, control and spread across populations. The objective is to generate key indicators for effective relevant clinical data clustering going to discover even weak signals, thus providing information for further reasoning.

2.3.1 *Analysis of social media content*

The analysis of big data can facilitate epidemic intelligence and disease surveillance (Eckmanns et al., 2019). In general, health data can be analysed for:

- Extracting health data pertaining to patient analysis;
- Monitoring behavioural and social health risks;
- Examining trends and patterns related to current diseases and drug responses;
- Screening of health and disease registries to improve efficiency;
- Other aspects of patient daily living can be considered to support precision medicine;
- Emotional and semantic data from call transcripts, survey responses, slangs, dialects.

In a crisis, key indicators can be generated for effective relevant clinical data clustering, going to discover weak signals, providing information for further reasoning. Solutions based on AI algorithms can gather contents available in the *surface, deep* and *dark web,* analysing them by deep semantic algorithms and discovering insights referred to specific health problems on a part of the population during a dedicated temporal reference in a geographic location (O'Shea, 2017).

2.3.2 *Event-based Internet bio surveillance*

A relatively new surveillance type has emerged called *event-based Internet bio surveillance systems* (O'Shea, 2017) able to use information on events impacting health

from online sources (e.g., social media or news aggregates). They are inexpensive, yet flexible and transparent, and search Internet sources for events that impacts health (SAS, 2021).

As healthcare organizations develop more sophisticated capabilities, they can move from basic descriptive, analytics towards predictive analytics and insights (Kandikonda, 2021). They can identify and extract details to assess risk, to apply prediction to estimate the likelihood of a future outcome based on trends and patterns. This approach facilitates prevention and improves health management. Such goal can be reached by leveraging a Natural Language Processing (NLP) (Linguamatics, 2021b,c) engine, enabling the extraction of information from raw unstructured data. Utilising rule-based and machine learning (ML) algorithms, data is loaded into a warehouse (SAS, 2021), where it is aligned and combined with existing structured data. This helps with the development of predictive and reasoning models that can be run against all the relevant features (Zhang et al., 2020).

2.4 Explainable artificial intelligence

Traceability and *explainability* are important components to support the decision-making processes that guide medical intelligence utilising explainable AI for ethics (Royal Society, 2019).

2.4.1 *The COGITO Hybrid Engine*

Expert.ai COGITO is a hybrid cross lingual cognitive engine based either on deep semantic technology (multilanguage ontology, going to unveil the real meaning of terms) or on machine/deep learning (ML/DL). It understands text in "the way people do" thus overcoming the need to have many tagged corpora for training pure ML/DL algorithms. The engine transforms unstructured Twitter *Real World Data* (RWD), by capturing contextualised meanings of words and concepts, into *actionable intelligence* (Expert.ai, 2021) for:

    a) Monitoring of behavioural health risks

    b) Patents' analysis

    c) Risk adjustments and disease registries

    d) Precision medicine

    e) Bio-surveillance algorithms

    f) Drug discovery, repurposing, and pharmacovigilance

    g) *Voice of the customer* (VoC).

In particular, the VoC data further expand, with insights from patients' forums, social media, survey answers, etc. (*ex-ante* analysis). Here, the focus is on understanding *emotions* (not only the sentiment) and the *style of writing (stylometric analysis)* about specific verbal groups, jargons, dialects, morphology on expressing opinions, to apply a reasoning on the social and psychological impact of a pandemic giving good information to doctors, psychologists and sociologists.

2.4.2 *Truthfulness and reliability of information*

Truthfulness and *reliability* of information are critical to inform any medical threat and safeguard the security of the general public (Royal Society, 2019). In the deep understanding of the information dynamics that lead decision-makers to trust or not on a reasoning and to consequently adopt it on specific missions, it is crucial to have real news and to be aware about fake ones (Ahmad et al., 2020; Aldwari and Alwahedi, 2018). This is achieved by measuring sites for consistency, spread, meaning underlying language, logic versus emotional contexts of data, comparing credible media sources and understanding social bots versus human web interactions. Specific morphology and packaging of messaging, mainly during web campaigns, can influence public perception and behaviour. In this case *stylometric analysis* and other semantic extractions can help to filter.

2.4.3 *Ethics and explainable AI*

According to the EU Commission's "Ethics Guidelines for Trustworthy AI", *Explainable Artificial Intelligence (*XAI) is still an open challenge for systems based on pure ML/DL algorithms and neural networks because parameters values resulting from training processes may be difficult to correlate with the outcomes (European Commission, 2020). Embedding ethical principles is a critical requirement to safeguard against bias (Expert Systems, 2021). In effect, COGITO is a XAI technology that can be configured, understood and supported by subject matter experts who can modify it by creating semantic rules or by applying machine learning or deep learning algorithms. This aspect is in line with *traceability* and *explainability,* going to ensure responsibility and accountability for AI outcomes.

3. The scenario

Digital disease detection and bio surveillance can be a great added value for early discovery of even weak signals coming from social media, mainly if the MEDINT platform is based on the International Press Telecommunications Council (IPTC) Taxonomy, the Medical Subject Heading (MESH) and the SNOMED CT clinical vocabulary readable by computers, well-recognized healthcare knowledge base, universally used by experts, practitioners, doctors.

3.1 Medical Intelligence Platform and GIS

3.1.1 *Interoperability of MIP*

The Expert AI Medical Intelligence Platform (MIP) (Figure 1) can be integrated with third party software such as automatic translation and SpeechToText engines for transcription from audio/video contents. Using its peculiar Application Programming Interface (API), the MIP could interoperate with different systems such as Command and Control (C2) or simulation systems. Furthermore, it can visualize, in its Graphical User Interface (GUI), different Geographic Information Systems (GIS) features (Gmap, Esri, etc.) for performing advanced geospatial analysis.
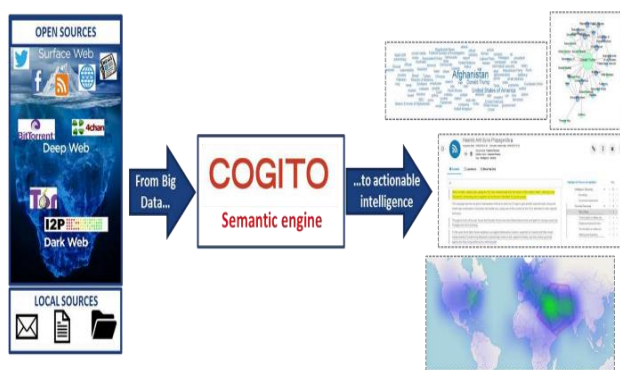
Figure 1. Medical Intelligence Platform (Expert.ai).

3.1.2 *Social media data in crisis management*

Social media data are currently used in several different domains, such as, for example, in disaster management, for generation of crisis maps, damages assessment, detection of help requests, lead of relief activities or provision of information (Schlosser et al., 2021).

In effect, in crisis management it is of the highest importance to have an avant-garde information-seeking capability as it reveals where data is transformed into information. We may want *to collect, discriminate and disseminate* relevant and reliable information in an intuitive way. Examining specific geographic regions, it is possible to discover even small signals, and understand medical threats among diverse racial and ethnic communities.

3.2 Tweets' geotagging

3.2.1 *Tweets geographic information*

In Twitter, the precise geotagging feature is off by default since 2019 but still users have the option to add their geographic location (latitude and longitude) information to their tweets. This allows Twitter to collect, store, and use their precise location, such as GPS information.

For users who have chosen to attach location information to their tweets, a selected location label is displayed underneath the text of their tweet. On twitter.com, the user can select a location label such as the name of a city or a neighbourhood (Twitter, 2021). With the Twitter in-app camera for iOS and Android, users can attach a photo or video to a tweet and toggle on the option to tag the precise location. In that case, the tweet will include both the location label chosen by the user and the geographic location (latitude and longitude) of his or her device. Photos and videos to confirm landmarks or specific locations can be utilised to improve the credibility of information shared.

However, for messages posted on Twitter, the geographic location (i.e., latitude and longitude) is available for only 1–2% of the messages (Twitter, 2021). Therefore, after collecting data, it is necessary to apply methods for automatically assigning a geographic location to each collected tweet (Schlosser et al., 2021), linking each text message to a specific area of a country. The level of geographical resolution is defined by the Nomenclature of Territorial Units for Statistics (NUTS-1) macro regions (https://ec.europa.eu/eurostat/web/nuts/background). Data available in social media platforms can be collected by (Liang and Zhu, 2017):

   a) Collection of "Log-data"

   b) Use of Application Programming Interfaces (APIs)

   c) Web scraping.

While web crawling can traverse only the public web, the use of APIs provides researchers with access to non-public internet environments (those requiring authentication through login and password) as the collection of data runs through the back-end of the social media service (Lomborg and Bechmann, 2014).

Authors have made use of the Twitter API. Twitter makes *"public Tweets and replies available to developers and allow developers to post Tweets via our API. Developers can access Tweets by searching for specific keywords or requesting a sample of Tweets from specific accounts. These endpoints are used by NGOs like the UN to identify, understand and counter misinformation around public health initiatives."* (Twitter, 2021).

Tweets' privacy depends on the IT structure where geo tweets and other data are loaded in. Therefore, compliance with the privacy rules have to be necessary managed by the IT Manager of the organization using the MIP tool. Obviously, it is possible to remove the geo tags from the MIP visual panel, but this works only at the visualization level.

In Figure 2, using MIP graphical widgets, authors created a boolean query, focusing the search on tweets in which users mention London geographical area AND the "breathing problems" concept. It is possible to filter thousands of tweets to have an overview of the most relevant ones in which occur words like "London", "pneumonia", "chest infection" and "can't breath" connected with the meaning of "breathing problems", regardless of whether in the communication channels.

Figure 2. Drawing a polygon in GeoMap

3.3 Mapping COVID-19 related tweets

3.3.1 *Scenario approximations*

In the context of this study, by authors' choice:

- Only tweets downloaded by using the Twitter free API (a little portion of the whole Twitter stream) have been ingested and processed;

- Only tweets written in English language have been considered.

3.3.2 *Mapping COVID-19 symptoms in London*

By targeting the London metropolitan area with MIP during the testing period, authors had the opportunity to receive in near-real time updated information on health situation regarding the *"COVID-19 symptoms in London"*

3.3.3 *Health, sentiment and emotions picture*

The Medical Intelligence Platform (MIP) provides, in near-real time, a health situational awareness regarding "COVID-19 symptoms in London" phenomenon.

In fact, the MIP has been tested to create an updated picture of *health, sentiment* and *emotion* situation among the population of the target region, as in Figure 3. In fact, it was possible to reach super-specialized tweets talking about *fear* due to COVID-19 symptoms in London region. Finally, it was also possible to monitor, in real-time, public opinion, sentiment and more than 70 different citizens' emotions such as, for example, *fear, anxiety* or *hope*. and

their impact on the local population.

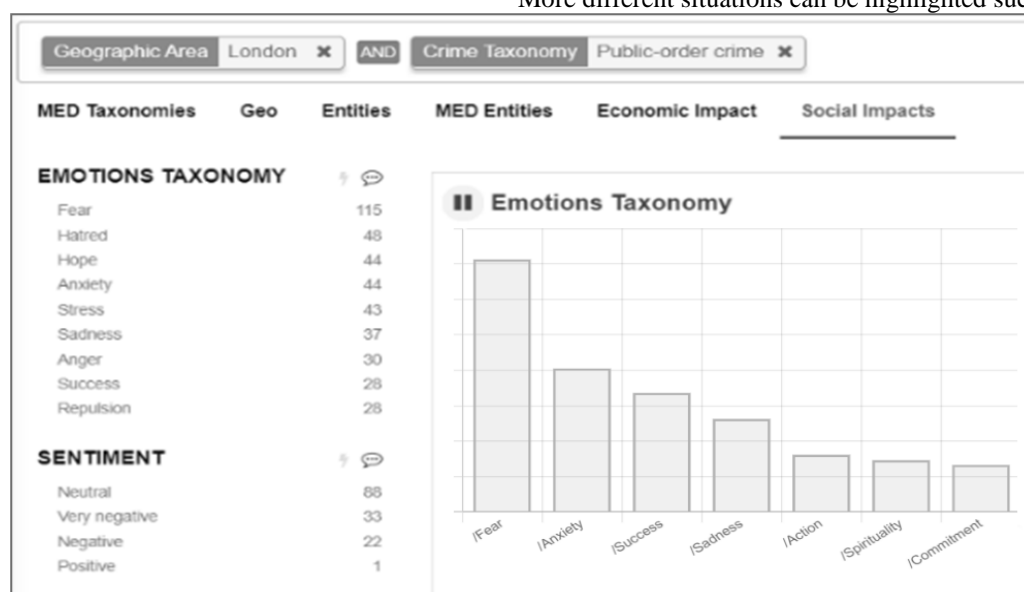More different situations can be highlighted such as, for

Figure 3. Emotions and sentiment analysis.

phenomenon. From here, authors could further analyse search results and manipulate search queries by using different perspectives and modules available in MIP. Analysts could also refine search results by submitting a geo-spatial query on the GeoMap module that integrate GIS features in MIP. GeoMap is not only a perspective with which is possible to have a "focused view" on the geographic area of interest, but it is also a powerful tool which let users draw a polygon and submit a geospatial query (see Figure 3).

Results could be analysed by using *domain-specific taxonomies* that contain, arranged in a hierarchically mode, a set of topics related to a specific domain. They could be used to:

- Understand how many documents are currently related to a specific topic (numbers rendered at right of each topic name),

- Further refine the search query adding a new constraint; by clicking on the name of a topic, the user will add a new constraint to the current search query.

example, an overview of the public order problems connected to COVID-19 disease in a selected geographic area.

Finally, it was also possible to create a dynamic graphical widget to have a multi-level overview of the COVID-19 phenomenon in the selected scenario.

The analyst can manipulate search queries and refine results by submitting a geo-spatial query on an area of interest, by utilising the GeoMap module that integrates Geographic Information Systems (GIS) features in MIP and generate statistics (Figure 4).
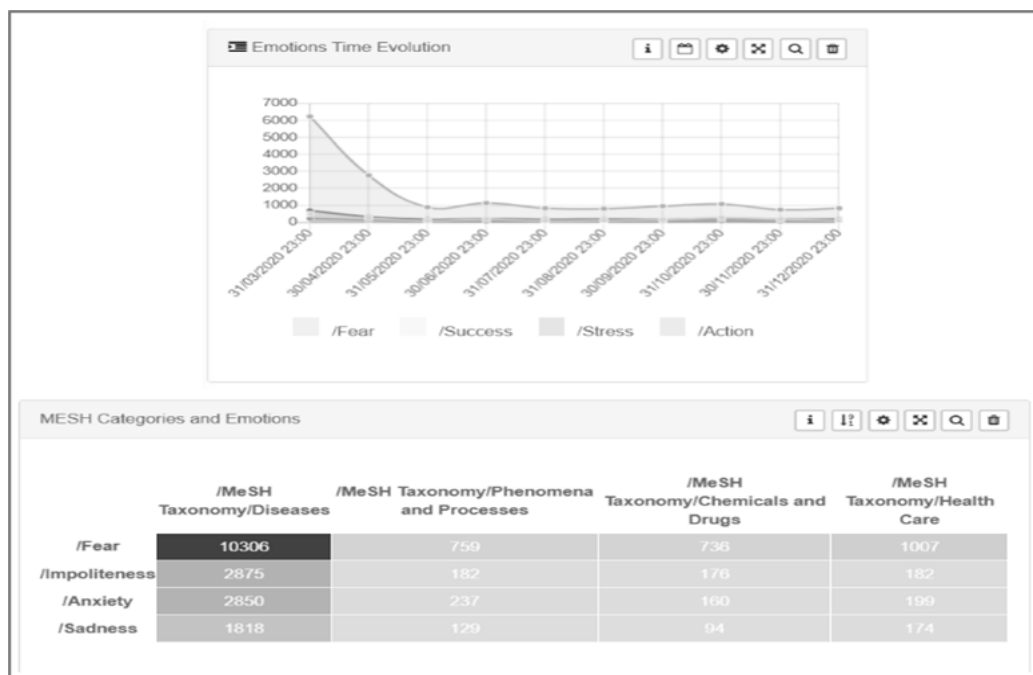
Figure 4. MIP generated statistics.

3.4 Results' discussion

In this study, authors tested an AI supported medical intelligence (MEDINT) tool to understand if it can provide better situational awareness to decision-makers, considering the scenario of London.

The use of a tool such as MIP, able to identify documents (news, tweets, etc.) incorporating multiple terms and phrases associated with "*breathing problems*" provided the chance to rapidly obtain accurate search results and to visualize an overview of the most relevant documents dealing about the selected "*breathing problems*" concept. The MIP platform proved to have the ability to: Acquire data from different type of sources (internal and open);

- Process information by making use of the Expert AI Cogito® advanced semantic engine;

- Present results in an effective way, by making use of interactive graphs, tabs and facet to rapidly and easily explore contents;

- Provide actionable insight and identifying risks and opportunities as data emerges.

The analysis supporting decision-making is facilitated by the AI and makes it possible to:

- *Understand: s*emantically index content, categorize it based on relevant domain taxonomies, identify events and entities;

- *Augment*: enable end users to process more information, faster, anticipate threats and

- acquire knowledge and insight from unstructured data;

- *Automate*: support all phases of the intelligence cycle, from source acquisition to deep analysis, and dissemination;

- *Monitor in real-time the target:* receive proactive alerts triggered by relevant news from validated sources; specific dashboards can provide a single view of target information.

4. Conclusions

The Expert.ai Medical Intelligence Platform (MIP), an explainable AI supported tool, has been tested for mapping and analytics of social media content generated by people in the context of the COVID-19 pandemic.

For this study, authors have chosen the Twitter microblogging service for its immense numbers of users, accessible volume of data and real-time data sharing and finally, for its consolidated role in disaster management and humanitarian action.

The medical intelligence (MEDINT) approach proved to be promising and instrumental to improve preparedness and planning in the context of medical threats.

Although there are ethical concerns regarding the explainability and the veracity of data, the MIP, in combination with embedding ethical principles, has demonstrated promising potential towards managing and processing large volumes of unstructured data, such as, for example, those contained in Twitter messages.

Such data can be transformed into comprehensible and interpretable information to guide and support decision-making processes. This is critical to national and international security and protection from impending and future catastrophic threats.

With no doubts, it is important to consider MEDINT to cover both prevention activities aiming to anticipate events (ex-ante analysis) and to analyse historical situations going to mitigate future events by using past experiences (ex-post analysis).

AI supported MEDINT can be deployed either at the *strategical level* or at *operational level,* in order to plan and conduct specific missions in a geographic location, with particular attention at ethics issues, as already exposed in the paragraph about explainable AI.

Results indicate that such approach, exploiting real-time social media data shared by people, provides a powerful and feasible solution for measuring and tracking the evolution of disease in society, for early detection of outbreaks and for supporting appropriate data driven decision-making.

5. References

Ahmad, I., Yousaf, M., Yousaf, S. and Ahmad, M.O. (2020). "Fake News Detection Using Machine Learning Ensemble Methods," *Complexity*, Oct. 17, 2020, Accessed Dec. 27, 2020, [Online]. Available: https://www.hindawi.com/journals/complexity/2020/8885861/

Aldwairi, M. and Alwahedi, A. (2018). "Detecting Fake News in Social Media Networks," *Procedia Computer Science*, vol. 141, pp. 215–222, doi: 10.1016/j.procs.2018.10.171.

Big Data Analytics in Healthcare. (2020). Accessed on: Dec. 27, 2020. [Online]. Available: https://www.linguamatics.com/solutions/big-data-analytics-healthcare.

Bowsher, G., Milner, C. Sullivan, R. (2016). "Medical intelligence, security and global health: the foundations of a new health agenda." J R Soc Med. Jul;109(7):269-73. doi: 10.1177/0141076816656483. PMID: 27385714; PMCID: PMC4941003.

Brownstein, J.S., Freifeld, C.C. and Madoff, L.C. (2009). "Digital Disease Detection - Harnessing the Web for Public Health Surveillance," *New England Journal of Medicine*, vol. 360, no. 21, pp. 2153–2157, doi: 10.1056/nejmp0900702. Accessed on Dec 27, 2020, [Online]. Available: https://www.nejm.org/doi/full/10.1056/NEJMp0900702

Curley, M.G. and Herrington, J. (2011). "The securitisation of avian influenza: international discourses and domestic politics in Asia." *Rev Int Stud*; 37: 141–166. Available: https://www.linguamatics.com/solutions/population-health-management-analytics

Dash, S.; Shakyawar, S.K., Sharma, M. *et al.* (2019). Big data in healthcare: management, analysis and future prospects. *J Big Data* 6, 54 https://doi.org/10.1186/s40537-019-0217-0.

Eckmanns, T., Füller, H. & Roberts, S.L. (2019). Digital epidemiology and global health security; an interdisciplinary conversation. *Life Sci Soc Policy* 15, 2, https://doi.org/10.1186/s40504-019-0091-8.

Eilandera, D., Trambauera, P., Wagemakerb, J., van Loenena, A. (2016). 12th International Conference on Hydroinformatics, HIC Harvesting social media for generation of near real-time flood maps, 1877-7058 © The Authors. Published by Elsevier Ltd

European Commission (2019). "Ethics guidelines for trustworthy AI," *European Commission*, Accessed Jan 14, 2020, [Online]. Available: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

Expert Systems (2018). Expert System Achieves a World First by Enhancing Different AI Capabilities Accelerating Human-Like Language Comprehension for Text Analytics. Accessed on Dec. 27, 2020, [Online]. Available: https://www.prnewswire.com/news-releases/expert-system-achieves-a-world-first-by-enhancing-different-ai-capabilities-accelerating-human-

like-language-comprehension-for-text-analytics-300647843.html, last accessed 2021/01/15.

Huang, X., Li, Z., Jiang, Y., Li, X., & Porter, D. (2020). Twitter reveals human mobility dynamics during the COVID-19 pandemic. PloS one, 15(11).

Huang, X., Li, Z., Jiang, Y., Ye, X., Deng, C., Zhang, J., & Li, X. (2021). The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the US during the COVID-19 pandemic. International Journal of Digital Earth, 1-19.

Kandikonda, S. (2020). "Artificial Intelligence and Data Science in Healthcare," Accessed Dec. 27, 2020, [Online]. Available: https://www.thedatasteps.com/post/data-science-in-healthcare.

Kaufman, D.D. (2001). Medical Intelligence: A Theatre Engagement Tool. Strategy Research Project. USA: Defense Intelligence Agency.

Kusumasari, B., Prabowo, N.P.A. (2020). Scraping social media data for disaster communication: how the pattern of Twitter users affects disasters in Asia and the Pacific. *Nat Hazards* **103,** 3415–3435 https://doi.org/10.1007/s11069-020-04136-z

Liang, H.; Zhu, J.J.H. (2017). Big Data, Collection of (Social Media, Harvesting). In The International Encyclopedia of Communication Research Methods; John Wiley & Sons, Inc.: Hoboken, NY, USA, pp. 1–18

Linguamatics (2021). Population Health Management and Analytics. Accessed on Dec 27, 2020, [Online]. Available: https://www.linguamatics.com/solutions/population-health-management-analytics Lomborg, S.; Bechmann, A. (2014). Using APIs for Data Collection on Social Media. Inf. Soc. 30, 256–265.

Murthy, Dhiraj. 2018. *Twitter.* Cambridge: Polity Press.

O'Shea, J. (2017). "Digital disease detection: A systematic review of event-based internet biosurveillance systems," *International Journal of Medical Informatics*, vol. 101, pp. 15–22, May 2017, doi: 10.1016/j.ijmedinf.2017.01.019.

Prieto, V.M., Matos, S., Álvarez, M., Cacheda, F. and Oliveira, J.L. (2014). "Twitter: A Good Place to Detect Health Conditions," *PLoS ONE*, vol. 9, no. 1, p. e86191, Jan. 2014, doi: 10.1371/journal.pone.0086191.

Reperant, L.A., Osterhaus, ADME. (2017). AIDS, Avian flu, SARS, MERS, Ebola, Zika… what next? Vaccine. 2017 Aug 16;35(35 Pt A):4470-4474. doi: 10.1016/j.vaccine. Epub PMID: 28633891; PMCID: PMC7115365.

SAS, https://www.sas.com/it_it/home.html, last accessed 2021/01/15.

Schlosser, S.; Toninelli, D.; Cameletti, M. (2021). Comparing Methods to Collect and Geolocate Tweets in Great Britain. J. Open Innov. Technol. Mark. Complex. 7, 44. https:// doi.org/10.3390/joitmc7010044

Seddighi, H., Salmani, I., Seddighi, S. (2020). Saving lives and changing minds with Twitter in disasters and pandemics: A Literature Review. Journal. Media. 1, 59-77.

Stylianou, A., Talias, M.A. (2017). Big data in healthcare: a discussion on the big challenges. *Health Technol.* **7,** 97–107 https://doi.org/10.1007/s12553-016- 0152-4.

The Royal Society (2019). *Explainable AI: the basics POLICY BRIEFING* Nov. 2019, Accessed: Jan. 14, 2021 [Online]. Available: https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf.

Twitter (2020). Accessed on Dec. 27, 2020. [Online]. Available: https://s22.q4cdn.com/826641620/files/doc_financials/2019/q1/Q1-2019-Slide-Presentation.pdf, last accessed 2021/01/14.

U.S. Department of the Army (1989). *Medical Intelligence in a Theater of Operations: Field Manual 8–10-8.* Accessed on Jan. 13, 2021. [Online]. Available: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1087&context=dodmilintel last accessed 2021/01/13.

White House Briefing (2014). U.S. Response to the Ebola Epidemic in West Africa Published: September 16, 2014. Accessed Jan. 13, 2021. Available: https://obamawhitehouse.archives.gov/the-press-office/2014/09/16/fact-sheet-us-response-Ebola-epidemic-west-africa.

Zhang, D., Yin, C., Zeng, J. *et al.* (2020). Combining structured and unstructured data for predictive models: a deep learning approach. *BMC Med Inform Decis Mak* 20**,** 280, https://doi.org/10.1186/s12911-020- 01297-6.