

# Leveraging OpenStreetMap for hyperlocal geocoding of Twitter data: A spatiotemporal analysis of the 2016 Haifa (Israel) wildfire

Motti Zohar<sup>1,\*</sup>, Bar Genossar<sup>2</sup>, Ronnen Avny<sup>3</sup>, Naama Tessler<sup>4</sup> and Avigdor Gal<sup>2</sup>

<sup>1</sup> School of Environmental Sciences, University of Haifa, Haifa, Israel; [motti.zohar@univ.haifa.ac.il](mailto:motti.zohar@univ.haifa.ac.il)

<sup>2</sup> Faculty of Data & Decision Sciences, Technion – Israel Institute of Technology, Haifa, Israel; [bargenossar@gmail.com](mailto:bargenossar@gmail.com); [avigal@technion.ac.il](mailto:avigal@technion.ac.il)

<sup>3</sup> Doctoral School of Business and Management, Corvinus University of Budapest, Budapest, Hungary; [ronnena@gmail.com](mailto:ronnena@gmail.com)

<sup>4</sup> Ecologist and Independent Consultant, Haifa, Israel; [naamates@gmail.com](mailto:naamates@gmail.com)

\* Author to whom correspondence should be addressed.

This abstract was accepted to the OSMScience 2025 Conference after peer-review.

The increasing frequency and severity of urban wildfires demand new sources of near real-time information to support emergency response and disaster risk reduction [1-4]. In this study, we present an application for extracting and georeferencing the spatiotemporal distribution of tweets associated with the November 2016 wildfire in Haifa, Israel. The unprecedented urban fire influenced densely populated neighborhoods, caused extensive infrastructural damage, and led to the evacuation of thousands of residents [5]. However, because of the nature and extent of the fire that lasted nearly 3 days, complete and reliable information concerning the emergence and development of new fire locations at the sub-city scale was only partial. Accordingly, management and decision-making procedures were complicated and some of the cascading events along with the occurrence of the catastrophe were hard to detect and addressed.

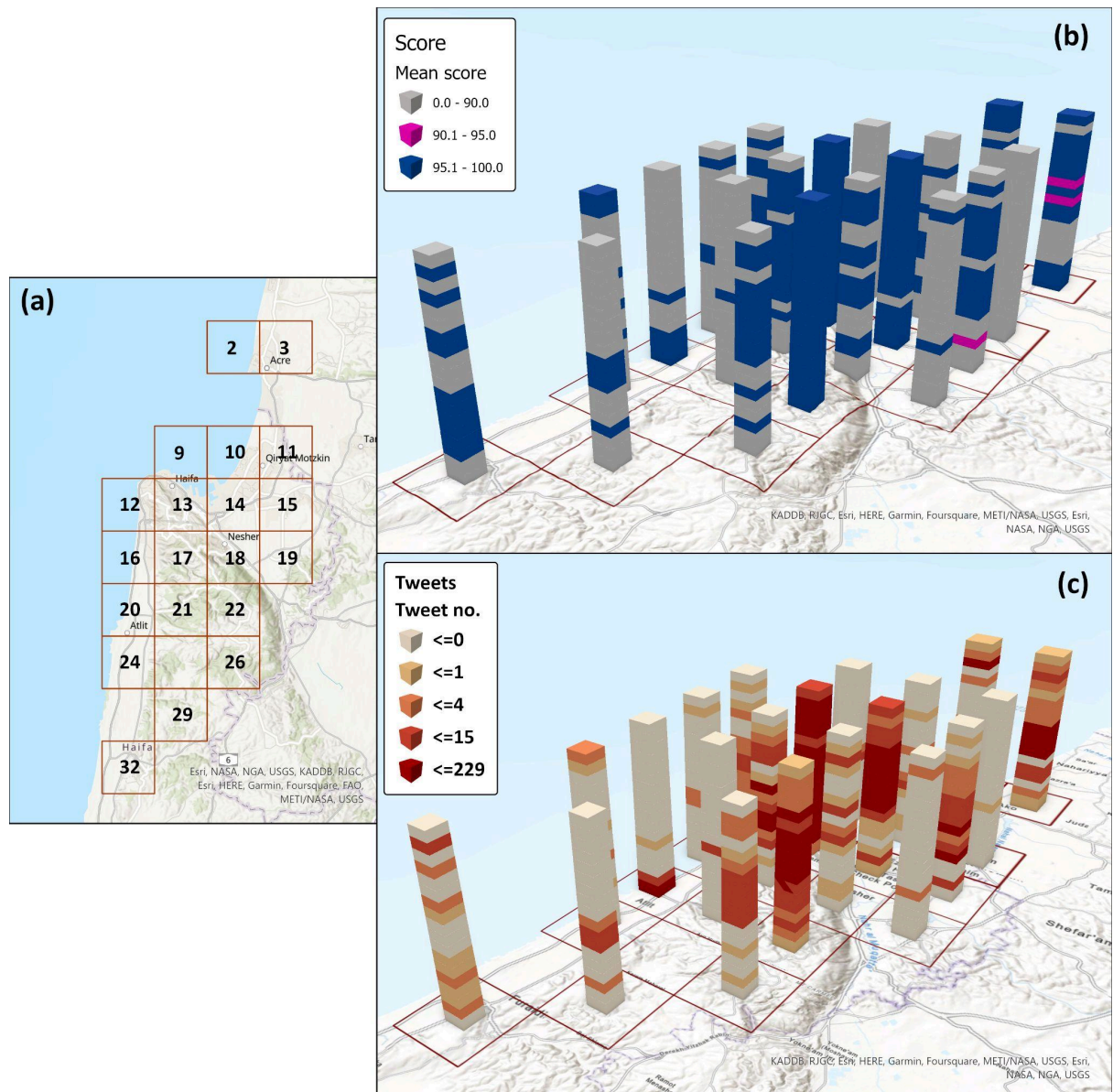
The purpose of the study was to analyze tweets as a potential source of near real-time information and examine to what degree Twitter can be used to assist decision-making during occurrences of urban catastrophes. The implemented research combined Natural Language Processing (NLP), Machine Learning (ML), and Geographic Information Science (GIScience) to filter, classify, and precisely geolocate tweets at the city, neighborhood or street-level resolution [6]. One of the main components of the established geospatial framework was OpenStreetMap (OSM) used in conjunction with the GeoNames gazetteer to construct a comprehensive spatial reference corpus. This enabled the geocoding of both explicitly and implicitly localized tweets that lack GPS metadata—an essential challenge given that only 1–3% of the tweets are geotagged with reliable geographic coordinates [7, 8].

We have purchased approximately 2.4 million tweets using keywords related to the wildfire (in the Hebrew, Arabic, and English languages) between November 24<sup>th</sup> –27<sup>th</sup>, 2016. After classification using topic modelling and RCNN (Recurrent Convolutional Neural



Networks) [9], around 114,000 tweets were labelled as relevant to the event. Of these, only 31 tweets were geotagged with geographic coordinates which is obviously an insufficient number of observations to perform spatial analysis. To overcome this shortcoming, we implemented a text-based georeferencing approach leveraging gazetteer data extracted from the OpenStreetMap and GeoNames databases. Accordingly, we converted 18 OSM shapefiles into a unified point dataset containing a wide variety of geographic features—ranging from neighborhoods and roads to public buildings and natural landmarks. This dataset was merged with a version of the GeoNames corpus to create a point-based localized gazetteer representing the Haifa metropolitan area and its environs. For the purpose of geocoding, each point was associated with its place name in Hebrew, Arabic, or English. Following, the geocoding pipeline consisted of the following key steps: (1) NLP techniques including tokenization, stemming, and stop-word removal to extract named entities and spatial references from the tweet's metadata [10]; (2) FuzzyWuzzy-based string matching algorithm that computes the Levenshtein distances between strings extracted from the tweet tokens and the place names in our gazetteer [11]; and (3) matchings were assigned a confidence score, which allowed us to weight the credibility and accuracy of the spatial data. For example, georeferenced tweets with scores above 90% were deemed highly reliable for trend detection.

The process yielded 16,672 georeferenced tweets distributed across 130 unique localities within Haifa and its close vicinity. To examine whether these locations of tweets represent fire hotspots, we added the official reported fire incidents by the Israel Fire and Rescue Services (IFRS) [12]. The geographic distributions of the georeferenced tweets as well as the IFRS data are presented in [13], accompanied with an online map [14]. Following, we conducted a spatiotemporal inspection by aggregating tweets into 5×5 km grid cells and 8-hour intervals—bins that were informed by the density of localities extracted from the OSM/GeoNames hybrid gazetteer. We used ESRI ArcGIS Pro to generate Kernel Density Estimation (KDE) maps [15] and 3D visualizations of the tweets' distribution. The 2D and 3D densities were then superimposed on the real IFRS data resulting in strong temporal and spatial correspondence. In most of the bins where actual fires were documented, relevant georeferenced tweets were also present. Additionally, tweets often captured the cascading nature of the fire, spreading across Haifa and adjacent towns. Importantly, the spatial granularity enabled us to detect not only urban hotspots but also peripheral areas such as the cities of Daliyat el-Carmel and Akko, which were underrepresented in the media reporting.



**Figure 1.** (a) the spatial distribution of grid-Cells in Haifa and environs; (b) Spatiotemporal distribution of georeferencing accuracy scores (results of Levenshtein distance algorithm used for token-locality matching) achieved to georeference tweets, ranging 0-90 and 99–100; (c) spatiotemporal distribution of aggregated tweets into separate 8-hour bins.

OpenStreetMap played a pivotal role in our ability to extract high-resolution geospatial insights from unstructured social media content. Its richness and multilingual tagging schema allowed effective token matching across Hebrew, Arabic, and English tweets. While GeoNames provided broad administrative and populated place names, OSM uniquely offered sub-city level details such as parks, fire stations, neighborhoods, and points of interest—that dramatically enhanced geocoding precision. The findings of the study demonstrate that OSM and GeoNames might function as an open, extensible backbone for disaster informatics, particularly in regions where official geospatial datasets are sparse or restricted. Additionally, this research showcases a replicable model for fusing crowd-sourced geographic data with user-generated content (e.g., Twitter) to inform

emergency responses at the hyperlocal scale. It is worth mentioning that since the conduction of the study, Twitter (X) has switched ownership and accordingly services it used to provide are prone to changes. Nevertheless, the study demonstrates principles and best practices of how crowd-sourced geographic data in conjunction with social media data might assist in acquiring near-real time information during crisis management.

## References

- [1] Kuligowski, E. (2021). Evacuation decision-making and behavior in wildfires: Past research, current challenges and a future research agenda. *Fire Safety Journal*, 120, 103129.
- [2] Castellnou, M., Prat-Guitart, N., Arilla, E., Larrañaga, A., Nebot, E., Castellarnau, X., Venderell, J., Pallàs, J., Herrera, J., Monturiol, M., Cespedes, J., Pagès, J., Gallardo, C., Miralles, M. (2019). Empowering strategic decision-making for wildfire management: avoiding the fear trap and creating a resilient landscape. *Fire ecology*, 15(1), 31.
- [3] Gao, J., Shalini, K., Gaur, N., Guan, X., Chen, S., Hong, J., & Mahmoud, M. (2017, August). Data-driven Forest fire analysis. In *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*, San Francisco, USA. pp. 1-7.
- [4] Desmond, M. (2011). Making firefighters deployable. *Qualitative Sociology*, 34(1), 59-77.
- [5] Michael, Y., Lensky, I. M., Brenner, S., Tchetchik, A., Tessler, N., & Helman, D. (2018). Economic assessment of fire damage to urban forest in the wildland–urban interface using planet satellites constellation images. *Remote Sensing*, 10(9), 1479.
- [6] Zohar, M. (2021). Geolocating tweets via spatial inspection of information inferred from tweet meta-fields. *International Journal of Applied Earth Observation and Geoinformation*, 105, 102593.
- [7] Huang, B., & Carley, K. M. (2019). A large-scale empirical study of geotagging behavior on Twitter. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining*, Vancouver, Canada. pp. 365-373.
- [8] Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., & Shook, E. (2013). Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday*, 18(5).
- [9] Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent convolutional neural networks for text classification. In *Proceedings of the AAAI conference on artificial intelligence* (V. 29, No. 1).
- [10] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- [11] Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8), 707-710.
- [12] Israel Fire and Rescue Services data of fire incidents in Haifa and environs 22-28 Nov 2016, [https://mzohar.haifa.ac.il/wp-content/uploads/2025/08/fire-haifa-2016\\_public.zip](https://mzohar.haifa.ac.il/wp-content/uploads/2025/08/fire-haifa-2016_public.zip)
- [13] Zohar, M., Genossar, B., Avny, R., Tessler, N., & Gal, A. (2023). Spatiotemporal analysis in high resolution of tweets associated with the November 2016 wildfire in Haifa (Israel). *International Journal of Disaster Risk Reduction*, 92, 103720.
- [14] Zohar, M. (2023). Online map: Spatiotemporal analysis of tweets associated with the November 2016 wildfire in Haifa. University of Haifa: Haifa, Israel, <https://arcg.is/1DmqmH>
- [15] Cai, X., Wu, Z., & Cheng, J. (2013). Using kernel density estimation to assess the spatial pattern of road density and its impact on landscape fragmentation. *International Journal of Geographical Information Science*, 27(2), 222-230.