Analysis of Canadian wildfire tweets over seven years

Sherritt, Braeden; Nejadgholi, Isar; Amini, Marzieh

Publisher's version / Version de l'éditeur:

Proceedings of the 38th Canadian Conference on Artificial Intelligence, 2025-05-19

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=f74410e3-9e20-425d-a81b-22d61eb0f2af
https://publications-cnrc.canada.ca/fra/voir/objet/?id=f74410e3-9e20-425d-a81b-22d61eb0f2af

Analysis of Canadian Wildfire Tweets Over Seven Years

Braeden Sherritt[†], Isar Nejadgholi[‡], Marzieh Amini[◇*]
[†] System and Computer Engineering department, Carleton University, Ottawa, Ontario, Canada
[‡] National Research Council Canada, Ottawa, Ontario, Canada
[◇] School of Information Technology, Carleton University, Ottawa, Ontario, Canada

Abstract

During wildfires, users share real-time updates, warnings, and personal experiences on social media, which offers valuable insights for emergency response and disaster management. However, the vast volume and unstructured nature of social media data pose challenges in effectively extracting meaningful information. The first Canadian-specific multimodal dataset for wildfire-related social media analysis, WildFireCan-MMD, was recently introduced, and a multimodal classifier was developed to classify social media posts into thirteen categories. In this study, we collect 46,279 posts from X, posted during the wildfire seasons (May–October) of 2018 to 2024, and label them using the trained classifier. We then analyze trends in wildfire-related discussions over seven years. Our findings reveal seasonal patterns in public discourse, with significant spike linked to heightened concerns over smoke and air quality. Analysis of wildfire season over a year uncovered a sequential progression in social media discussions: an initial rise in reports of firefighting efforts was followed by increased posts about evacuations and emergency updates, with a delayed peak in smoke and air quality concerns as wildfire smoke spread. These insights show how social media captures the dynamic nature of wildfire events, reflecting public awareness and response as disasters unfold. Our study demonstrates the value of automated classification in extracting actionable intelligence from large-scale social media data.

Keywords: Disaster Response, Canadian Wildfire Dataset, Multimodal Classifier, Social Media Analysis

1. Introduction

Social networks can serve as valuable sources of real-time information during natural disasters, such as wildfires. Users can share first-hand experiences, information, and updates on the situation that unfolds around them, making these platforms a valuable repository of information for disaster response teams. However, the volume and noisy nature of this content provide challenges in extracting the relevant information from the background. The need for an automated approach to analyze and then classify these wildfire-related posts has become evident. Machine learning tools and specifically multimodal classification models can be an instrumental tool in achieving this goal.

Canada's most destructive natural disaster is wildfires, which occur almost every year in the western provinces of Alberta and British Columbia [1]. This is why it is important for Canada to develop systems that can mitigate this damage as much as possible, especially with the increase in global warming. However, automatic triaging of social media data during wildfires in Canada is underexplored, and datasets surrounding this task are scarce.

2. Background and Literature Review

In a bibliometric analysis by [2], the use of social media in disaster management was systematically reviewed. This analysis highlighted key publications in the field and proposed future research directions, such as integrating big data analytics for more effective response efforts. Similarly, [3] focused on the public health implications of social media usage during

[†] braedensherritt@cmail.carleton.ca [‡]isar.nejadgholi@nrc-cnrc.gc.ca [◇*] marzieh.amini@carleton.ca

emergencies, emphasizing its effectiveness in spreading emergency warnings and assessing public needs. In addition to these studies, [4] provided a systematic review of deep learning applications in disaster response, discussing both the successes and challenges of these technologies. The review also outlined guidelines for future research, stressing the importance of using multimodal data. Reference [5] further explored how social media analytics can improve disaster response by enabling the rapid gathering of time-sensitive information from affected areas. Reference [6] aimed to improve disaster relief efforts by analyzing social media data to better understand public attitudes and the demand for relief supplies across various disaster types. Reference [7] introduced a machine learning model that utilized multimodal data, including satellite imagery and social media, to predict housing and shelter needs in the aftermath of disasters.

The lack of labeled social media imagery data for disaster response has also been a significant challenge. Reference [8] addressed this issue by creating CrisisMMD, a multimodal dataset derived from Twitter that covers various natural disasters. CrisisMMD is a well-known natural disaster dataset collected during seven major events, including earthquakes, hurricanes, wildfires, and floods that happened in the year 2017 across different parts of the world. The authors of [8] expanded upon their work by creating a model to classify this data automatically, using a combination of BERT and CNN structures [9]. Expanding upon the work by [8] and [9], [10] created a model using the CrisisMMD dataset, updated with the use of transformers. However, this dataset does not include any data from Canada.

In [11], a taxonomy is developed to categorize Canadian wildfire-related content, and WildFireCan-MMD is introduced, the first Canadian-specific multimodal dataset tailored for wildfire events. This dataset comprises 4,688 user-generated posts from X (formerly Twitter) during the wildfires of 2022- 2024, each manually labeled according to the defined taxonomy. Furthermore, a multimodal classifier was designed and trained, capable of categorizing these posts and enabling the structured analysis of wildfire-related social media content.

3. Methodology

Building upon [11], we have significantly expanded the dataset by collecting an additional 46,279 posts from May to October in the years 2018 to 2024. Instead of manually labeling this new dataset, which is both time-consuming and resource-intensive process, we leveraged the previously trained classifier to categorize these posts. This newly labeled dataset allows for a broader analysis of the wildfire-related discussions and trends on social media and how users share information during such crises. This section presents the methodology for collecting and classifying the expanded dataset, followed by an analysis of trends observed across wildfire seasons.

For data collection, we used a similar methodology to [11], where we used hashtags based on the locations of active wildfire areas per year in question. Since our goal is to analyze the dataset and we do not aim to use it for testing classifiers, we do not remove the overlap between this dataset and the ones collected in [11]. To classify the newly collected dataset, we employed the previously trained multimodal model by [11], which consists of a text processing head using RoBERTa and an image processing head using ViT. This classifier achieved an F1-weighted average score of 83% on the test set of WildFireCan-MMD, demonstrating its effectiveness in categorizing wildfire-related posts. RoBERTa was chosen as the text classification head due to its effectiveness on NLP tasks as found in the literature [12]. It often achieves high accuracy in text classification and other downstream tasks due to its more robust pretraining. For the image classification head, ViT was chosen for its ability to capture long-range dependencies in images using self-attention mechanisms. Unlike traditional CNNs, which rely on localized feature extraction, ViT models images as sequences of patches, enabling it to recognize complex patterns more effectively [13].

Each post in the expanded dataset was fed into the classifier, which processed text and images to generate predictions. The model assigned a category to each post based on the predefined taxonomy, allowing structured labeling of the dataset. This automated classification approach allowed for efficiency in processing the large volume of newly collected data without having to manually label each individual sample. This, of course, will not uphold the same gold standard of labeling that a human annotator could achieve, and manual labeling will have to be done in the future. Individual classifications should be approached with skepticism, but this method should still allow for general trends in the data to be analyzed on a broader scale. The labeled dataset was then analyzed to identify trends and patterns in discussions related to wildfires over time. By examining variations in category distribution across different wildfire seasons, we aimed to gain insights into how social media users respond to wildfires. The previously defined taxonomy by [11] is shown in Table 1.

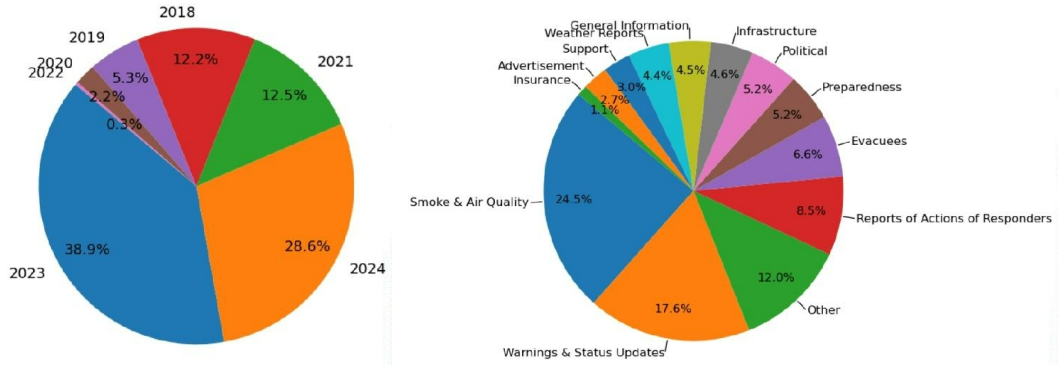
Table 1. Categories of WildFireCan-MMD, and their descriptions.

Category	Description
Evacuees	Information relating to evacuees, their movements, needs, and location.
Smoke and Air Quality	Tweets related to or showing signs of smoke or the current air quality.
General Information	General info about the wildfire situation, such as total hectares burned.
Preparedness	Information for the general public to prepare themselves and their property.
Weather Reports	Information relating to the weather with a specific location mentioned.
Warnings & Status Updates	Fire bans in certain areas, new information about a specific area, updates from officials.
Reports of Actions of Responders	Actions of responders within specific areas or times, including prescribed burns.
Infrastructure	Detours, road closures, damage to infrastructure (e.g., utility poles, highways), repairs by crews.
Political	Posts directed towards political figures or parties (excluding situation updates).
Support	Mental health and financial support, temporary housing for livestock.
Insurance	Information relating to insurance, employment insurance, and EI benefits.
Advertisement	Posts about food, restaurants, off-topic ads for services (e.g., apps, air purifiers, not insurance-related).
Other	No 'useful' information, focusing on images (e.g., scenery), general complaining, or irrelevant content.

4. Experimental Results & Discussion

Figure 1 shows the general statistics of the collected data, showing most of this data belongs to years 2023 and 2024 and categories of 'Smoke and Air Quality' and 'Warning and Status Updates'. To better understand the temporal distribution of wildfire-related discussions, we analyzed the total volume of posts collected per week, as illustrated in Figure 2. Since all posts were collected between May and October, we overlaid the data from different years onto the same plot. This revealed that wildfire discussions typically begin increasing in the Spring as users warn others about the upcoming wildfire season. The highest peak in activity occurs between mid-July and mid-August, coinciding with the peak wildfire season. Interestingly, a significant spike is also observed in early June.

To investigate the cause of this early June peak, we examined the data on a per-year basis, as shown in Figure 3. This revealed that the anomaly originated from the year 2023.



(a) Ratio of data coming from each year. *(b)* Distribution of predictions on dataset.

Figure 1. General statistics of the collected dataset

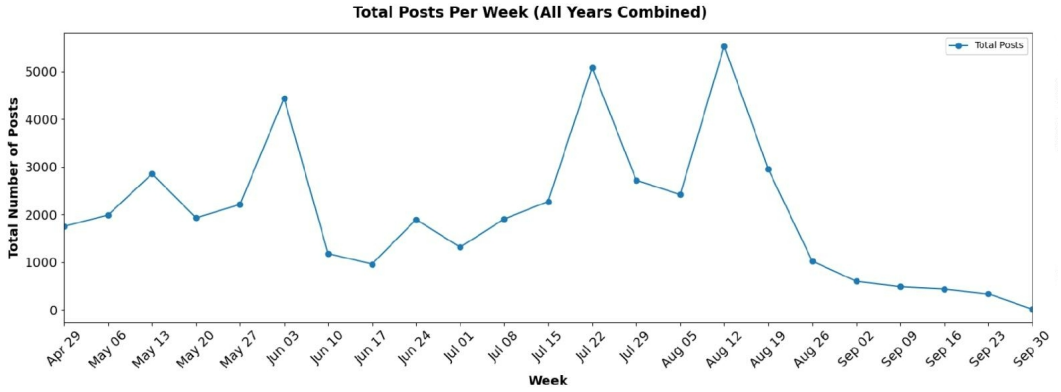


Figure 2. Total number of posts per week.

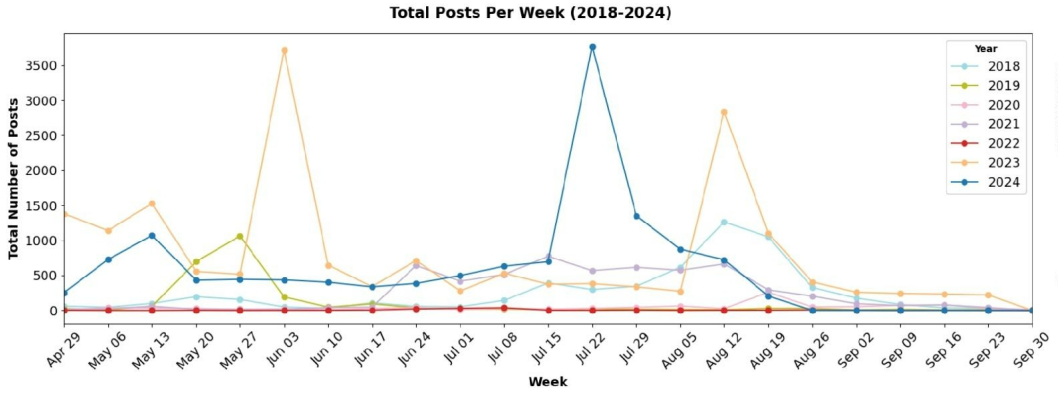


Figure 3. Total number of posts per week sorted by year.

Further analysis, presented in Figure 4, categorized the posts from this period into their respective classes. The results showed that during the week of June 1st, nearly 2,500 posts were labeled as containing references to smoke and air quality concerns. This suggests that a major wildfire event or a significant deterioration in air quality occurred during this time, prompting widespread social media discussions. The same can be seen around the

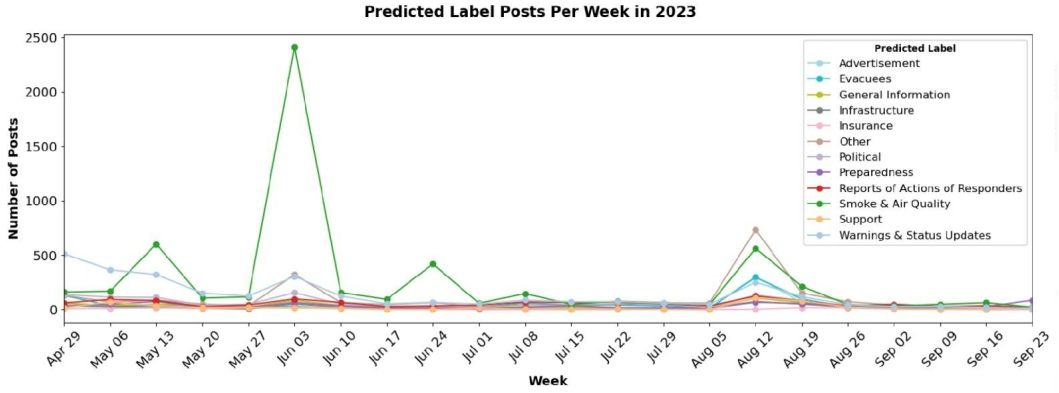


Figure 4. Total number of posts per week in 2023 sorted by category.

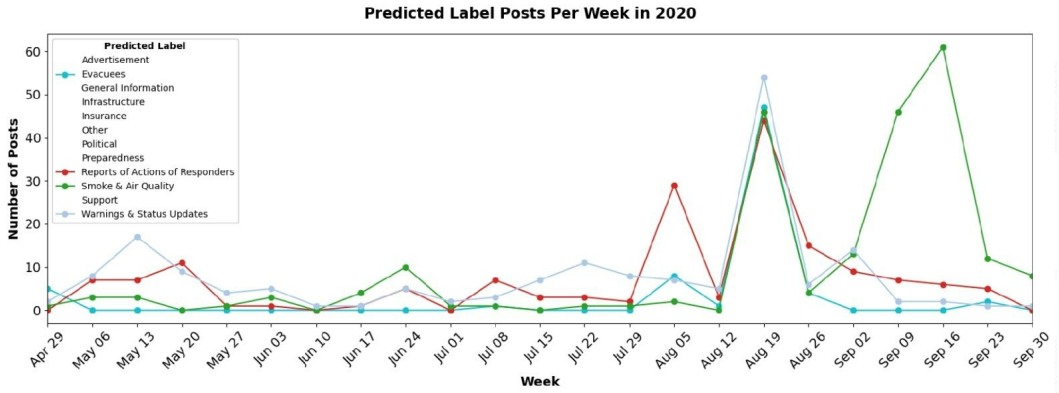


Figure 5. Total number of posts per week in 2020 for 4 important categories.

week of August 12, where a major spike in all categories can be seen, most importantly the 'Evacuees' and 'Warning & Status Updates' categories, indicating that during major wildfire events social media is used in Canada to communicate such information.

Further analysis of the data from 2020 uncovered another interesting trend, which can be seen in Figure 5. By filtering out some of the most critical categories, we observed a pattern in the distribution of the posts. Around the week of August 5, there was a noticeable increase in the number of posts categorized as 'Reports of Actions of Responders'. Two weeks later, this was followed by a significant rise in posts related to 'Evacuees' and the other mentioned categories. This suggests a sequential progression in wildfire response and public engagement, where initial reports of firefighting efforts are followed by discussions regarding evacuations and emergency updates. Over the subsequent month, an increase in 'Smoke and Air Quality' posts was observed, likely reflecting the widespread impact of lingering wildfire smoke. This pattern highlights how social media can capture the evolving stages of wildfire events, from emergency response to broader environmental effects. Contextual explanation for the events of 2020 in relation to Figure 5 can be found in Appendix A.

5. **Conclusion**

In this study, we collected over 46,000 Canadian wildfire-related social media posts spanning several years and labeled them using a previously developed multimodal classifier,

integrating RoBERTa and ViT models. We then analyzed trends in wildfire discussions on social media over seven years. Our findings show the evolving nature of public engagement with wildfire events, revealing seasonal patterns and response dynamics on social media activity. The insights gained from this research further highlight the potential of social media as a valuable tool for disaster response and public awareness.

Acknowledgements

We thank the National Program Office of the National Research Council (NRC) Canada for funding this project. This work was also supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] C. F. Service. *Canadian National Fire Database – Agency Fire Data*. Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre, Edmonton, Alberta. 2025. URL: https://cwfis.cfs.nrcan.gc.ca/ha/nfdb.

[2] M. A. Fauzi. "Social media in disaster management: review of the literature and future trends through bibliometric analysis". In: *Natural Hazards* 118 (2 2023). ISSN: 15730840. DOI: 10.1007/s11069-023-06079-7.

[3] K. Muniz-Rodriguez, S. K. Ofori, L. C. Bayliss, J. S. Schwind, K. Diallo, M. Liu, J. Yin, G. Chowell, and I. C.-H. Fung. "Social media use in emergency response to natural disasters: a systematic review with a public health perspective". In: *Disaster medicine and public health preparedness* 14.1 (2020), pp. 139–149.

[4] N. Algiriyage, R. Prasanna, K. Stock, E. E. H. Doyle, and D. Johnston. "Multi-source Multimodal Data and Deep Learning for Disaster Response: A Systematic Review". In: *SN Computer Science* 3 (1 2022). ISSN: 2662-995X. DOI: 10.1007/s42979-021-00971-4.

[5] T. Acikara, B. Xia, T. Yigitcanlar, and C. Hon. "Contribution of social media analytics to disaster response effectiveness: A systematic review of the literature". In: *Sustainability (Switzerland)* 15.11 (2023), p. 8860.

[6] Z. S. Dong, L. Meng, L. Christenson, and L. Fulton. "Social media information sharing for natural disaster response". In: *Natural Hazards* 107 (3 2021). ISSN: 15730840. DOI: 10.1007/s11069-021-04528-9.

[7] K. S. Ochoa and T. Comes. "A Machine learning approach for rapid disaster response based on multi-modal data. The case of housing & shelter needs". In: *arXiv preprint arXiv:2108.00887* (2021).

[8] F. Alam, F. Ofli, and M. Imran. "CrisisMMD: Multimodal twitter datasets from natural disasters". In: *12th International AAAI Conference on Web and Social Media, ICWSM 2018*. 2018. DOI: 10.1609/icwsm.v12i1.14983.

[9] F. Ofli, F. Alam, and M. Imran. "Analysis of social media data using multimodal deep learning for disaster response". In: *Proceedings of the International ISCRAM Conference*. Vol. 2020-May. 2020.

[10] R. Koshy and S. Elango. "Multimodal tweet classification in disaster response systems using transformer-based bidirectional attention model". In: *Neural Computing and Applications* 35 (2 2023). ISSN: 14333058. DOI: 10.1007/s00521-022-07790-5.

[11] B. Sherritt, I. Nejadgholi, and M. Amini. *WildFireCan-MMD: A Multimodal dataset for Classification of User-generated Content During Wildfires in Canada*. 2025. arXiv: 2504.13231 [cs.CV]. URL: https://arxiv.org/abs/2504.13231.

[12] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. 2019. arXiv: 1907.11692 [cs.CL]. URL: https://arxiv.org/abs/1907.11692.

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. 2021. arXiv: 2010.11929 [cs.CV]. URL: https://arxiv.org/abs/2010.11929.

Appendix A. Broader Context of the Analysis in 2020

In the following, we provide a summary of the report provided by the Canadain Interagency Forest Fire Center[1], for the 2020 wildfire season, which expalins our observations in Figure 5. The 2020 fire season in Canada was relatively quiet, with significantly fewer fires and less area burned than average, but it was framed by intense activity abroad. Early in the year, Canada supported Australia's record-breaking fire season by sending personnel to Victoria, New South Wales, and South Australia until mid-March. As COVID-19 emerged, Canadian fire agencies had to adapt operations for safety. Domestically, the fire season began slowly, with the first interagency support request in late May coming from New Brunswick, rather than the usual western provinces. Through June and July, fire activity escalated in Québec, which received support from Manitoba, Ontario, Alberta, and Newfoundland and Labrador. Ontario and Labrador also saw significant fires and required air support. By August, severe drought in the Maritime provinces led to more fire activity, especially in New Brunswick and Nova Scotia. As the Canadian season wound down, resources shifted to the United States, where over 590 Canadian personnel helped combat devastating wildfires in California and Oregon from late August to mid-October. Using this information, we can deduce that the spike observed in August was due to fire activity in the Maritime provinces, and the rise in discussion around 'Smoke & Air Quality' was due to the events happening south of the border, where Canadian personnel assisted with wildfires in the western United States.

[1] https://www.ciffc.ca/sites/default/files/2021-02/Canada_Report_2020_Final.pdf