DETECTION OF LOCAL CRISIS EVENTS: A CASE STUDY ON COLORADO WILDFIRES USING SOCIAL MEDIA AND SATELLITE IMAGERY

Marwen Bouabid and Mohamed Farah

Univ. Manouba, ENSI, RIADI LR99ES26, University Campus Manouba, 2010, Tunisia
email: marwen.bouabid@ensi-uma.tn
Univ. Manouba, ISAMM, University Campus Manouba, 2010, Tunisia
email: mohamed.farah@riadi.rnu.tn

ABSTRACT

Social media has become vital for disseminating natural disaster information, but its reliance has limitations. This study underscores the crucial integration of diverse data sources, particularly remote sensing and social sensing, to overcome challenges faced by traditional uni-modal methods. Using the Colorado wildfire as a case study, we applied advanced techniques, including named entity recognition and natural language processing, to extract tweet locations, grouping them based on geographical relevance. We extracted images from tweet URLs to enrich satellite imagery, associating them with relevant tweet groups. This research combines theoretical and practical insights, addressing detection challenges within a simplified framework. It highlights the innovative integration of social media and satellite imagery, offering valuable disaster response and management directions. This research aims to clarify the link between social media data and remote sensing data for researchers. Specifically, it aims to enrich satellite imagery with images and text from tweets to provide a clearer view of events and their locations.

***Index Terms*—** Wild Fire Detection, Social Media, Remote Sensing, linking social and remote sensing, change detection, Machine learning, classification.

1. INTRODUCTION

Social bookmarking platforms like Twitter, Facebook, and YouTube play a pivotal role in enabling swift event sharing. They allow users to convey thoughts succinctly through geotagged content, including photos and live videos. Particularly valuable in crises, users share real-time, location-specific information, expediting the dissemination of critical data.

In this context, our study emphasizes the essential integration of diverse data sources, specifically social media and remote sensing, to address disaster detection and monitoring challenges. Focusing on the Colorado wildfire as a case study, advanced techniques, including named entity recognition and natural language processing, were employed to extract tweet locations. These tweets were grouped based on geographical relevance, and associated images from tweet URLs were integrated to enrich the satellite imagery dataset.

This research combines theoretical and practical insights, addressing detection challenges within a simplified framework. The study highlights the innovative integration of social media and satellite imagery, offering valuable disaster response and management directions. The amalgamation of these diverse datasets significantly enhances the accuracy and efficiency of local crisis event detection. Social media datasets from projects like CrisisLex and CrisisNLP and specific time and geo-coordinates of events sourced from platforms like Wikipedia for remote sensing applications are invaluable resources. This wealth of data facilitates focused testing of approaches, streamlining the research process in both domains and their collaborative fusion.

2. RELATED WORK

Event detection, originating from the Topic Detection and Tracking (TDT) project, involves identifying temporal and spatial topics in data sources. It is categorized into Retrospective Event Detection (RED) or New Event Detection (NED), emphasizing comprehensive information collection, including event description, time, and location. Takahashi et al. [1] explored Twitter's utility in emergencies, considering user time and location. Local events, natural occurrences at specific locations and times, are a focal point of research. Quezada et al. [2] extract geo-located events from social bookmarking websites.

Satellite imagery is pivotal for disaster monitoring, employing change detection techniques for applications like target detection, regional planning, and warfare. Studies explore satellite data for disaster identification, mapping, and crisis prediction, involving pixel-based and object-based change detection processes [3] [4] [5].

The fusion of social media and satellite imagery addresses

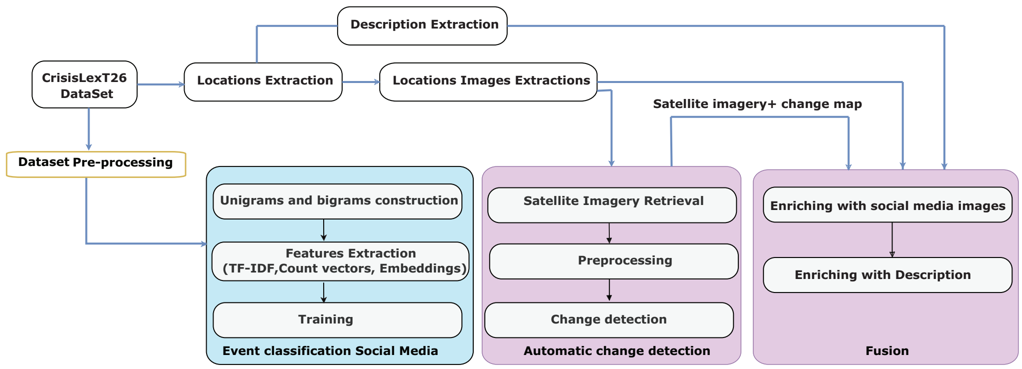

Fig. 1: Practical Framework for Combining Social and remote sensing

limitations. Systems like Jord, introduced by Kashif et al. [6], collect social media data on natural disasters, linking it with remotely sensed data. Other approaches, such as CrisMap [7] and GeoSensor [8], enhance change detection over satellite images with event content from social media. The multi-source data framework proposed by Xu et al. [9] integrates Weibo data, RS images, and historical geographic information for waterlogging probability assessment.

Innovative methodologies, like Wang et al. [10]'s fusion of social media and remote sensing data for weather-driven hazard prediction, showcase the potential of this integration in time-critical applications. Rapid flood inundation mapping by Rosser et al. [11] combines geotagged photos from social media, optical remote sensing, and high-resolution terrain mapping. The Selective Attention Mechanism (SAM) approach proposed by Bouabid et al. [12] links social media and remote sensing in crises.

3. THE TASK OF LOCAL CRISIS EVENT DETECTION

Studying crisis events in social media and remote sensing requires careful data analysis. Integrating diverse sources like social media and satellite imagery poses challenges, considering data scarcity and resolution issues. The task's complexity requires nuanced approaches, especially when integrating social media and satellite imagery [13]. Focusing on the 2012 Colorado wildfires, including the Waldo Canyon Fire in Colorado Springs and the High Park Fire near Fort Collins, this study addresses the complexities of event detection. We aim to provide insights into the challenges of combining social media and satellite imagery.

4. A PRACTICAL EXAMPLE FOR LINKING SOCIAL MEDIA AND REMOTE SENSING

We integrated theoretical insights with practical knowledge to address the challenge of local crisis event detection, utilizing a framework illustrated in Figure 1. This framework was applied to the crisisLExT26 dataset, linked with satellite imagery based on event-specific dates and locations. To enhance the dataset's completeness, we supplemented it with information from Wikipedia, which provided precise event times and locations. To detect events on social media, we trained classifiers such as Support Vector Machine, Logistic Regression, Random Forest, Gradient Boosting, Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). Given the diverse nature of events, we approached the detection as a change detection problem for remote sensing.

4.1. Event classification in social Media

Approaching local crisis event detection as a retrospective RED classification problem, we utilized reliable text classification classifiers: Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Gradient Boost (GB), Gated Recurrent Unit (GRU), and Long Term Short Term Memory (LSTM).

The CrisisLexT26 dataset [14], comprising 26 events of various types, was our target. We reduced similar and replicated classes, resulting in 14 classes while retaining their respective tweets. The tweets were then split into unigrams and bigrams (n-grams with n=1,2), followed by the construction of Count Vectors, TF-IDF Vectors, and word embeddings. Subsequently, we trained these algorithms for local suspicious event classification and evaluated their performance.

4.1.1. Dataset and Pre-processing

This dataset includes 250k tweets from 26 local crisis incidents in 2012 and 2013 (2K-4K tweets per event), with the following Key attributes: Tweet ID, Text, Source, Type, and Informativeness. The analysis focused on Text and Informativeness. It covers diverse language tweets, including retweets, URLs, usernames, and emojis.

Tweets labeled "Not related" or "Not applicable" are "Off-Topic." "Related and Informative" or "Related but not Informative" tweets are labeled with local crisis event names. The dataset lacks events related to protests or riots.

Fig. 2: Examples of images showing Fire

Pre-processing used NLTK and WordNet to filter non-English tweets and remove stop words, punctuation, URLs, and usernames. Stemming with SnowballStemmer handled word morphological variations.

4.2. change detection

This section focuses on local crisis event detection from satellite imagery, treating it as an automatic change detection problem. Due to data limitations in the crisislexT26 dataset, unsupervised change detection is employed on multispectral images, analyzing two multi-temporal satellite images to identify changes between timestamps. Landsat Imagery, retrieved from the USGS Earth Explorer, involves band joining, radiometric corrections, addressing Landsat 7 scanline errors, and subtracting post-event imagery from pre-event imagery. The resulting pixel-by-pixel subtraction highlights the event, expressed mathematically.

$$diff(i, j) = |Image1(i, j) - Image2(i, j)| \qquad (1)$$

Using PCA and K-means clustering, we distinguished changing areas from unchanged ones in the crisislexT26 dataset. Minimizing the time slot between selected imagery dates was crucial for effective change detection. We aimed to enhance social media data with remote sensing imagery for a comprehensive global perspective of diverse events.

4.3. Linking Social and Remote sensing

Leveraging tweet text for image extraction enhances disaster event understanding. Satellite imagery offers a bird's-eye view, while social media images provide on-the-ground context. Using advanced techniques like named entity recognition and natural language processing, we associate images from tweet URLs with specific tweet groups, enriching the satellite imagery dataset. This combined approach offers a comprehensive view of spatial dynamics, facilitating a closer examination of specific locations. Tweet text is crucial in providing descriptive information about casualties, damage assessment, and other critical aspects. This holistic view aids authorities in informed decision-making during crisis response and recovery efforts.

4.3.1. Location and images extraction

Utilizing NLP and Named Entity Recognition (NER) is crucial for extracting precise geographic locations from crisis-related tweets. NLP analyzes tweet text to identify named entities representing locations. Tweets with the same geographic context are grouped, and associated images from tweet URLs are integrated, refining spatial granularity and enriching the dataset with on-the-ground visual insights. This fusion enhances location-based tweet grouping and provides contextually relevant images, aiding decision-makers in crisis response efforts for improved situational awareness. Figure 2 shows images with fires, and Figure 3 depicts images with no fires.

4.3.2. Textual content Extraction

After grouping tweets by location, NLP is crucial for extracting detailed textual descriptions, focusing on damage assessment, fire extent, casualties, and relevant details. Systematically analyzing the textual data in each tweet group aims to provide a comprehensive and nuanced description of the disaster event. NLP aids in distilling crucial information from real-time social media discourse, contributing to a detailed narrative for better overall comprehension. This assists authorities and organizations in more effective decision-making and response planning.

5. RESULTS

5.1. Tweets classification

Classification results, presented in with metrics like precision, recall, and f1-measure, are further detailed in Table 1. Logistic Regression with count vectors excels, achieving an accuracy of 93.21%. Support Vector Machine (SVM) follows closely with an accuracy of 92.82% using count vectors, while Long Term Short Term Memory (LSTM) scores 92.57%. Logistic Regression and SVM perform better with count vectors, while Random Forest and Gradient Boost excel with TF-IDF vectors. Gated Recurrent Unit (GRU) demonstrates comparable results, with accuracies of 91.67%, 90.91%, and 91.70%, respectively.

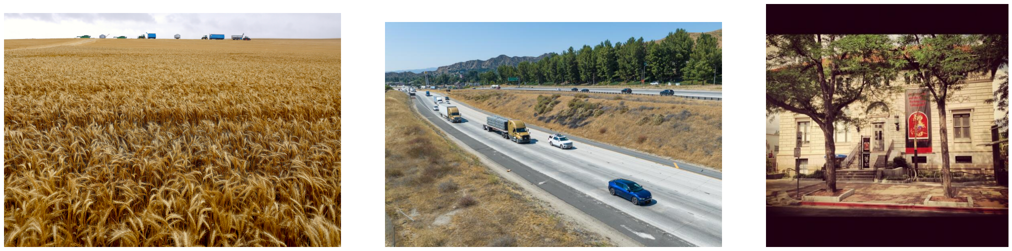

Fig. 3: Examples of images with No Fire

Table 1: Models and baselines performance

Model	Accuracy	F1 Score	Precision	Recall
SVM, Count Vectors	92.82	94	94	94
SVM, TF-IDF Vectors	91.72	93	84	88.27
LR, Count Vectors	93.21	94	94	94
LR, TF-IDF Vectors	88.99	93	92	92.49
RF, Count Vectors	92.04	92	93	92
RF, TF-IDF Vectors	91.67	93	92	92.49
GB, Count Vectors	91.58	93	94	93
GB, TF-IDF Vectors	90.91	93	92	92.49
LSTM, Embeddings	92.57	92	93	92
GRU, Embeddings	91.70	92	93	93

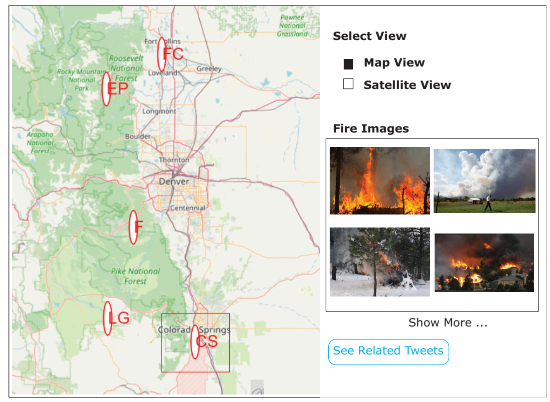

Fig. 4: Fusion Outcomes

5.2. Change Detection outcomes

Due to the unavailability of multiple sensor types for the crisisLexT26 dataset dates, we opted for Landsat Multispectral imagery. Employing automated change detection on images associated with dataset events, our approach effectively detected large-scale events such as fires. However, additional sensors were needed for small-scale events, which could have been more practical given their timing. Multispectral imagery has limitations, offering resolutions of 30 and 15 meters (pansharpened). Figure 5 illustrates the fire in the Blue Mountain Area in Australia, showing the change map relative to the 2012 Australian fire using pre-event Landsat 8 and post-event Landsat 7 images.

Despite efforts to remove scanline errors, complete removal was not achieved due to preprocessing performed before removal. We focused on promptly identifying events rather than achieving pinpoint accuracy in detecting changes.

5.3. Fusion outcomes

Fusing tweets and satellite imagery is a multifaceted process designed to understand disaster events comprehensively. After obtaining satellite imagery capturing the events, we integrate this data with dynamic change maps to visualize the extent of the fire and its evolving impact. To enrich this satellite imagery, we concurrently display on-the-ground images associated with specific locations extracted from social media. This integration enhances the spatial context and provides a more detailed and context-rich perspective. Furthermore, the fusion involves the display of selected tweets containing valuable information about the event, such as descriptions of specific locations, damage assessments, and casualty reports. Combining these varied sources of satellite imagery, on-the-ground social media images, and informative tweets, the fusion process creates a more holistic representation of the disaster, empowering decision-makers with a comprehensive view of effective response coordination and management. Figure 4 illustrates the outcomes.

6. CONCLUSION

This paper aims to leverage theoretical and practical knowledge for local crisis event detection, emphasizing the significance of location extraction in linking social media and remote sensing. The approach addresses multimodal classification and change detection within the crisislexT26 dataset, categorized into 14 classes based on event types. Classifier evaluation employs precision, recall, f1-measure, and accuracy metrics. Automatic change detection techniques are utilized, including the differencing image technique, PCA, and K-means. Results indicate approximate performance. The paper proposes a framework for enriching satellite data with textual descriptions and ground social media imageries, prioritizing speed over accuracy in detecting events of various types, considering the diverse range of events in local crisis detection with varying scales, particularly noting the limita-

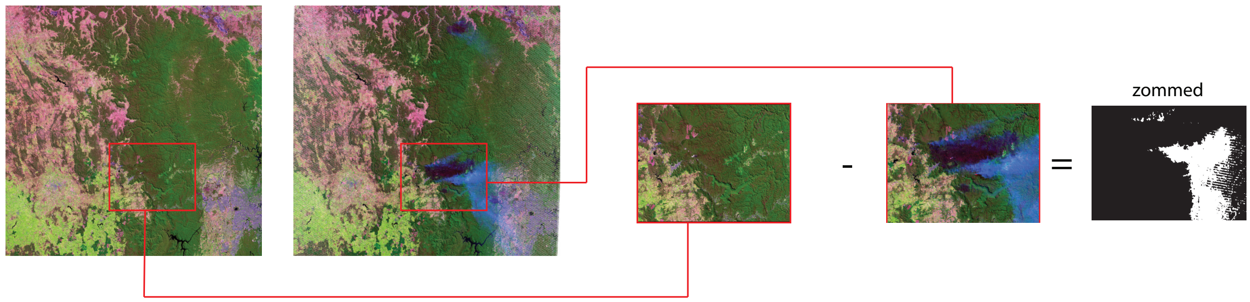

Fig. 5: Change Detection

tion in detecting small-scale events due to Landsat imagery resolution.

7. REFERENCES

[1] B. Takahashi, E. C. Tandoc Jr, and C. Carmichael, "Communicating on twitter during a disaster: An analysis of tweets during typhoon haiyan in the philippines," *Computers in Human Behavior*, vol. 50, pp. 392–398, 2015.

[2] M. Quezada, V. Peña-Araya, and B. Poblete, "Location-aware model for news events in social media," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2015, pp. 935–938.

[3] X. Huang, C. Wang, and Z. Li, "A near real-time flood-mapping approach by integrating social media and post-event satellite imagery," *Annals of GIS*, vol. 24, no. 2, pp. 113–123, 2018.

[4] Q. D. Cao and Y. Choe, "Building damage annotation on post-hurricane satellite imagery based on convolutional neural networks," *Natural Hazards*, vol. 103, no. 3, pp. 3357–3376, 2020.

[5] K. Thangavel, D. Spiller, R. Sabatini, S. Amici, S. T. Sasidharan, H. Fayek, and P. Marzocca, "Autonomous satellite wildfire detection using hyperspectral imagery and neural networks: A case study on australian wildfire," *Remote Sensing*, vol. 15, no. 3, p. 720, 2023.

[6] A. Kashif, K. Pogorelov, M. Riegler, N. Conci, and P. Halvorsen, "Social media and satellites," *Multimedia Tools and Applications*, vol. 78, no. 3, pp. 2837–2875, 2019.

[7] M. Avvenuti, S. Cresci, F. Del Vigna, T. Fagni, and M. Tesconi, "Crismap: a big data crisis mapping system based on damage detection and geoparsing," *Information Systems Frontiers*, vol. 20, no. 5, pp. 993–1011, 2018.

[8] N. Pittaras, G. Papadakis, G. Stamoulis, G. Argyriou, E. K. Taniskidou, E. Thanos, G. Giannakopoulos, L. Tsekouras, and M. Koubarakis, "Geosensor: semantifying change and event detection over big data," in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*. ACM, 2019, pp. 2259–2266.

[9] L. Xu and A. Ma, "Coarse-to-fine waterlogging probability assessment based on remote sensing image and social media data," *Geo-spatial Information Science*, pp. 1–23, 2020.

[10] H. Wang, E. Skau, H. Krim, and G. Cervone, "Fusing heterogeneous data: A case for remote sensing and social media," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 12, pp. 6956–6968, 2018.

[11] J. F. Rosser, D. Leibovici, and M. Jackson, "Rapid flood inundation mapping using social media, remote sensing and topographic data," *Natural Hazards*, vol. 87, no. 1, pp. 103–120, 2017.

[12] M. Bouabid and M. Farah, "Crisis detection by social and remote sensing fusion: A selective attention approach," in *International Conference on Computational Collective Intelligence*. Springer, 2023, pp. 350–362.

[13] B. Marwen, F. Mohamed, and F. I. Riadh, "Suspicious local event detection in social media and remote sensing: Towards a geosocial dataset construction," in *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. IEEE, 2020, pp. 1–6.

[14] A. Olteanu, C. Castillo, F. Diaz, and S. Vieweg, "Crisislex: A lexicon for collecting and filtering microblogged communications in crises," in *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.