# Human-Sensors & Physics aware Machine Learning for Wildfire Detection and Nowcasting

Jake Lever[1,2,3][0000−0002−1888−8699], Sibo Cheng[1,2][0000−0002−8707−2589], and Rossella Arcucci[1,2,3][0000−0002−9471−0585]

[1] Data Science Institute, Imperial College London, UK
[2] Leverhulme Centre for Wildfires, Environment & Society, UK
[3] Department of Earth Science and Engineering, Imperial College London, UK

**Abstract.** This paper proposes a wildfire prediction model, using machine learning, social media and geophysical data sources to predict wildfire instances and characteristics with high accuracy. We use social media as a predictor of wildfire ignition, and a machine learning based reduced order model as a fire spead predictor. We incorporate social media data into wildfire instance prediction and modelling, as well as leveraging reduced order modelling methods to accelerate wildfire prediction and subsequent disaster response effectiveness.

**Keywords:** Wildfires · Data Science · Machine Learning.

## 1 Introduction

Real-time forecasting of wildfire dynamics has attracted increasing attention recently in fire safety science. Twitter and other social media platforms are increasingly being used as real-time human-sensor networks during natural disasters, detecting, tracking and documenting events [13]. Current wildfire models currently largely omit social media data, representing a shortcoming in these models, as valuable and timely information is transmitted via this channel [14]. Rather, these models use other data sources, mainly satellites, to track the ignition and subsequently model the progression of these events. This data can often be incomplete or take an infeasible amount of preprocessing time or computation. Subsequently, the computation of current wildfire models is extremely challenging due to the complexities of the physical models and the geographical features [8]. Running physics-based simulations for large-scale wildfires can be computationally difficult. The combination of these factors often makes wildfire modelling computationally infeasible, due to both the unavailability or delay in the data, and the computational complexity of the subsequent modelling.

We show that by including social data as a real-time data source, setting up a Twitter based human sensor network for just the first days of a massive wildfire event can predict the ignition point to a high degree of accuracy. We also propose a novel algorithm scheme, which combines reduced-order modelling (ROM) and recurrent neural networks (RNN) for real-time forecasting/monitoring of

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_33

the burned area. An operating cellular automata (CA) simulator is used to compute a data-driven surrogate model for forecasting fire diffusions. A long-short-term-memory (LSTM) neural network is used to build sequence-to-sequence predictions following the simulation results projected/encoded in a reduced-order latent space. This represents a fully coupled wildfire predictive & modelling framework. The performance of the proposed algorithm has been tested in a recent massive wildfire event in California - The Chimney Fire.

In using social data coupled with a fast and efficient model, we aim to help disaster managers make more informed, socially driven decisions. We implement machine learning in a wildfire prediction model, using social media and geophysical data sources to predict wildfire instances and characteristics with high accuracy. We demonstrate that social media is a predictor of wildfire ignition, and present aforementioned novel modelling methods which accurately simulate these attributes. This work contributes to the development of more socially conscious wildfire models, by incorporating social media data into wildfire instance prediction and modelling, as well as leveraging a ROM to accelerate wildfire prediction and subsequent disaster response effectiveness.

## 2 Background & Literature Review

Sentiment analysis has been used in several studies [16, 3, 4] to predict natural disasters. These studies analyse social media content to detect sentiments related to natural disasters and identify potential warnings or updates. [15] used Twitter data and sentiment analysis to identify tweets related to wildfires and classify them based on the level of urgency. The study utilized natural language processing techniques to extract features from the tweets, such as location, hashtags, and keywords, and trained a machine learning model to classify the tweets into categories such as warnings, updates, or irrelevant. In addition to sentiment analysis, some studies have also used other techniques such as image analysis and environmental data integration to improve the accuracy of wildfire detection [18, 12]. For example, [23] used a combination of machine learning models and environmental data such as temperature and humidity to predict the occurrence and spread of wildfires accurately. In summary, sentiment analysis has shown promise to predict and detect wildfires. By analyzing social media content and identifying relevant sentiments, researchers can improve the efficiency and accuracy of real-time detection systems which is crucial for real-time wildfire nowcasting. Machine learning (ML)-based reduced order surrogate models [17, 5] have become a popular tool in geoscience for efficiently simulating and predicting the behavior of complex physical systems. Advantages of these models include their ability to effectively capture the nonlinear relationships and high-dimensional input-output mappings in geoscience problems and their ability to operate in real-time [6]. These models can be easily combined with data assimilation techniques to perform real-time corrections [9]. Additionally, machine learning algorithms can be trained on large amounts of data, making it possible to effectively incorporate a wide range of observations and simulations into the

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_33

modeling process [8]. ML-based ROMs have shown promise in the application of wildfire forecasting [8, 24, 25]. These models can effectively capture the complex relationships between inputs such as weather patterns, topography, and vegetation, and outputs such as fire spread and intensity. Advantages of these models include the ability to operate in real-time, the ability to effectively incorporate a wide range of observations and simulations, and the ability to incorporate the effects of time-varying variables such as wind speed and direction.
In this paper, we apply a surrogate model based on offline cellular automata (CA) simulations [2] in a given ecoregion. The physics-based CA model takes into account the impact of geophysical variables such as vegetation density and slope elevation on fire spread. The surrogate model consists of POD (Proper Orthogonal Decomposition) [19] (also known as Principle Component Analysis (PCA)) and LSTM (Long Short-Term Memory) neural network [20]. Wildfire spread dynamics are often chaotic and non-linear. In such cases, the use of advanced forecasting models, such as LSTM networks, can be highly beneficial in providing accurate and timely predictions of fire spread.

## 3 Methods

The pipeline of the proposed Human-Sensors & Physics aware machine learning framework is shown in Figure 1. The main two component consist of the ignition point prediction using the Human Sensor network and the fire spread prediction using a ML based ROM.

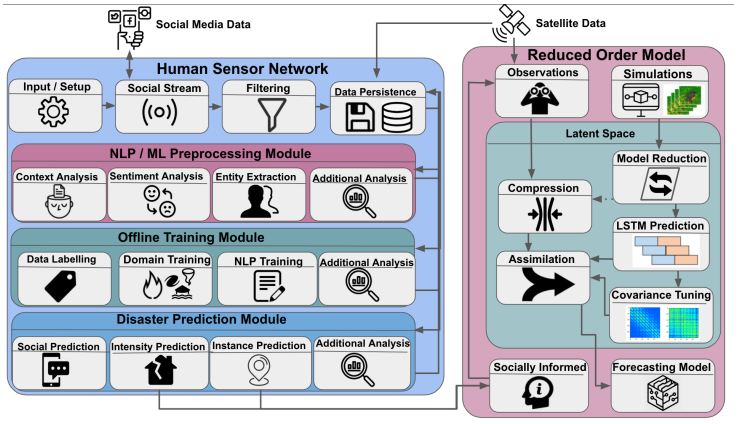


**Fig. 1.** Workflow of the proposed Human-Sensors & Physics aware Machine Learning framework

*Ignition point prediction using social media:* The data used for the instance prediction stage of the coupled model is taken from Twitter, using an academic licence for the V2 API, allowing for a full archive search of all tweets

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_33

from the period of the event. The twitter data queried was geotagged with a place name, which was subsequently geocoded to generate coordinates for the place of posting. The historical tweets were queried using the following query: '$(ChimneyORFireORWildfireOR\#ChimneyFire)place : Californiahas : geo'$, meaning the query was for the first keywords, posted in the region of California, and containing geolocation data. The query was run on a full historical search from public accounts from 12th-13th August 2016 - the first day of the wildfire event. For this period, in this location, 154 Tweets were downloaded and analysed. The resulting tweets then formed the dataset for the prediction. These tweets were analysed using a BERT transformer model for disaster classification, using the same methodology as used in [1]. This model analyses the text content of the message, and allocates disaster related labels based on the content of the post. Only posts with labels, i.e. classified by the model to be disaster related, were considered. Following this, the place name was extracted from the filtered Tweets via the Tweet metadata. The place name is given by a $place_id$, a hash code allocated by the API. Some of these also contained coordinates, but those that didn't were geocoded using the API to generate $lon$ and $lat$ coordinates for the tweet. Finally, all of the filtered, disaster related Tweets for the first day of the event had been allocated coordinates. Following this, named entity extraction was performed using google Entity Analysis API. This API uses NLP to extract named entities from a string of text. For this stage of the analysis, only entities with the type 'LOCATION' were extracted. These entities extracted were then geocoded using the Google Maps Geocoding API. Finally, the coordinate list for the tweet location and named entity location were averaged and used to make the final prediction for the wildfire ignition point. The averages were performed with a higher weighting on the locations extracted from the Entity Analysis, as these had been shown to be more indicative of accurate wildfire reporting.

*Physics Aware Wildfire Nowcasting* The ignition point computed in the first part, is here used as input to predict the fire spread. The physics simulation is implemented using a CA model, a mathematical model used for simulating the behavior of complex systems [2, 21]. In this study, the basic idea behind CA is to divide the landscape into a grid of cells, each of which can be in a certain state (e.g., unburned, burning, burned) [2, 8]. At each time step, the state of a cell is updated based on the states of its neighboring cells and a set of rules that determine how fire spreads. To generate the training dataset of machine learning algorithms, we perform offline CA simulations with random ignition points in a given ecoregion. To do so, an ecoregion of dimension $10 \times 10$ km$^2$ (split equally to $128 \times 128$ cells for CA) is chosen. 40 fire events with random ignition points are generated using the state-of-the-art CA model [2]. We then train a ML model using the simulations as training data. The ML model can then be used for unseen fire events in the ecoregion. A reduced space is then computed implementing a POD, a mathematical method used in dynamical systems to reduce the complexity and dimensionality of large-scale systems. It represents the system's behavior using a reduced set of orthogonal modes. These modes are obtained through a singular value decomposition of the time-series data as shown in many applica-

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_33

tions [10, 7, 26]. In this study, POD is applied to perform dimension reduction for wildfire burned area, where 100 principle components are used to describe the whole system. More precisely, the temporal-spatial snapshots, obtained from offline CA simulations [2] are collected to compute the principle components of the burned area. These offline CA simulations are carried out with randomly selected ignition point in the ecoregion. This ensures the good generalizability of the proposed approach when dealing with unseen data. Finally a ML model is trained in the reduced space using a LSTM [20], a type of recurrent neural network commonly used in machine learning and deep learning applications. In a reduced latent space, LSTMs are used to model complex sequential data in a compact and computationally efficient manner [11, 22, 5]. After the dimension reduction, by working in the reduced latent space, the LSTM network is able to handle the high-dimensional input data. In this work,we train a sequence-to-sequence LSTM model for burned area forecasting. The model take 4 snapshots (equivalent of 24 hours of fire spread) to predict the next 4 time steps of fire propagation. To form the training datasets for both POD and LSTM, 40 simulations with random ignition points are created. Each simulation is performed with 4 days of fire spread after ignition. Once the prediction in the lo.w-dimensional latent space is performed, the predicted latent vector can be decompressed to the full physical space. Forecasting the spread of a wildfire at the beginning of the fire is crucial because it helps decision-makers allocate resources effectively and prioritize evacuation plans.

## 4 Results: wildfire ignition and spread prediction

We tested the proposed approach on the Chimney wildfire event. The system has analysed twitted data as detailed in Section 3 and the ignition point result in (-119.72232658516701, 37.53677532750952) as (latitude, longtitude) coordinates. This point has been used as starting point for the fire spread prediction detailed later. The predicted fire spread of the first four days is illustrated in Figure 2. The image background represent the associated vegetation distribution in the ecoregion where the green color refers to a high vegetation density. The dimension of the ecoregion is about $10 \times 10\text{km}^2$. As explained in Section 3, CA simulations are carried out offline *a priori* to generate training data for the ML surrogate model. The developed ML model then manages to handle unseen ignition scenarios with the closest ecoregion. It can be clearly observed from Figure 2 that the wildfire prediction model exhibits a high level of performance, with a simulated fire spread that demonstrates a strong alignment with relevant geological features, namely the vegetation distribution. In fact, as described in Section 3, the ML model learns the fire-vegetation relationship from the CA simulation. It is clearly shown in Figure 2 that the area with higher vegetation density will have a higher probability to be burned out. In this study, we focus on the initial phase of fire spread (i.e., the first four days) since providing in-time fire spread nowcasting at the beginning of the fire event is crucial for fire fighting decision making. We also compare the predicted evolution of the burned

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_33

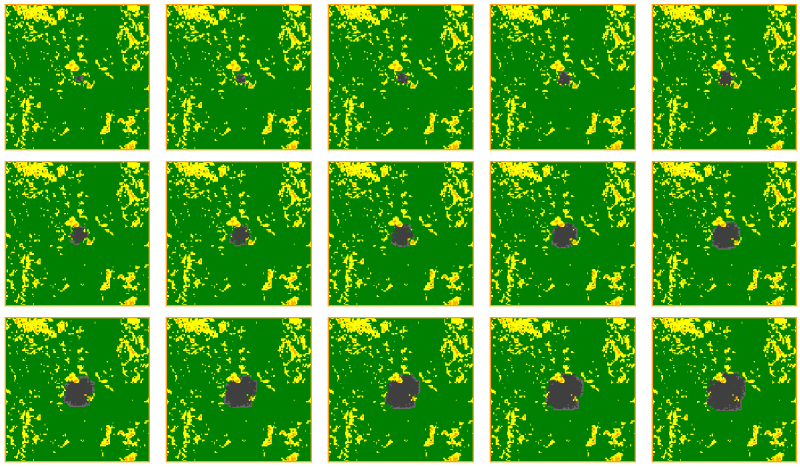

**Fig. 2.** Predicted sequences of burned area with the ignition point predicted by human sensors. The time interval is 6 hours and the first four time steps are provided by the CA simulation as the initial input of the ML model.

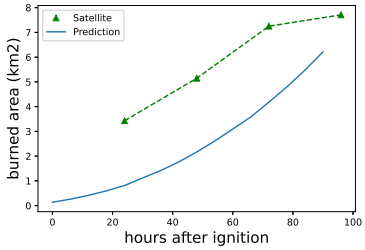


**Fig. 3.** The burned area after ignition: Satellite vs. ML prediction

area in $km^2$ to the satellite observations in Figure 3. It can be clearly seen that the predicted burned area in $km^2$ exhibits a growth trajectory similar to that of the observed values, which demonstrates the robustness of the proposed approach. The comparison between the proposed POD + LSTM approach and the original CA simulations has been performed in [8] for recent massive wildfires in California.

## 5 Conclusion & Future Work

The human-sensors & physics aware machine learning proposed in this paper provide an end-to-end framework that provides reliable fire ignition detection/localization and early-stage fire spread nowcasting. By including social data as a real-time data source, we show that setting up a Twitter based human sensor network for just the first days of a massive wildfire event can accurately predict

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_33

the ignition point of the event. Subsequently, using our novel algorithm, we show that the predicted burned area for each day of the event can be accurately modelled quickly and efficiently without using conventional data sources. The combination of using this real-time data source and a ROM system, we propose a lightweight coupled framework for real time wildfire detection and modelling. This work employs and develops the concept of the human-sensor in the context of wildfires, using users' Tweets as noisy subjective sentimental accounts of current localised conditions. By leveraging this social data, the models make predictions on wildfire instances. Subsequently, these instances are modelled using a fast, computationally efficient and accurate wildfire prediction model which is able to predict ignition points and burned area. We found that the main error in the prediction of fire ignition was that the ignition point prediction was biased towards more highly by populated areas. This result is to be expected to an extent, as there would naturally be more viewers and therefore sensors of an event in these locations. To combat this, a future work can aim to improve the methodology by taking into account the population density.

## References

1. Disaster tweets classification in disaster response using bidirectional encoder representations from transformer (bert). IOP Conference Series: Materials Science and Engineering **1115**(1), 012032 (mar 2021)
2. Alexandridis, A., Vakalis, D., Siettos, C., Bafas, G.: A cellular automata model for forest fire spread prediction: The case of the wildfire that swept through spetses island in 1990. Applied Mathematics and Computation **204**(1), 191–201 (2008)
3. Bai, H., Yu, G.: A weibo-based approach to disaster informatics: incidents monitor in post-disaster situation via weibo text negative sentiment analysis. Natural Hazards **83**(2), 1177–1196 (2016)
4. Beigi, G., Hu, X., Maciejewski, R., Liu, H.: An overview of sentiment analysis in social media and its applications in disaster relief. Sentiment analysis and ontology engineering: An environment of computational intelligence pp. 313–340 (2016)
5. Cheng, S., Chen, J., Anastasiou, C., Angeli, P., Matar, O.K., Guo, Y.K., Pain, C.C., Arcucci, R.: Generalised latent assimilation in heterogeneous reduced spaces with machine learning surrogate models. Journal of Scientific Computing **94**(1), 1–37 (2023)
6. Cheng, S., Jin, Y., Harrison, S.P., Quilodrán-Casas, C., Prentice, I.C., Guo, Y.K., Arcucci, R.: Parameter flexible wildfire prediction using machine learning techniques: Forward and inverse modelling. Remote Sensing **14**(13), 3228 (2022)
7. Cheng, S., Lucor, D., Argaud, J.P.: Observation data compression for variational assimilation of dynamical systems. Journal of computational science **53**, 101405 (2021)
8. Cheng, S., Prentice, I.C., Huang, Y., Jin, Y., Guo, Y.K., Arcucci, R.: Data-driven surrogate model with latent data assimilation: Application to wildfire forecasting. Journal of Computational Physics p. 111302 (2022)
9. Cheng, S., Quilodrán-Casas, C., Ouala, S., Farchi, A., Liu, C., Tandeo, P., Fablet, R., Lucor, D., Iooss, B., Brajard, J., et al.: Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review. arXiv preprint arXiv:2303.10462 (2023)

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_33

10. Gong, H., Cheng, S., Chen, Z., Li, Q., Quilodrán-Casas, C., Xiao, D., Arcucci, R.: An efficient digital twin based on machine learning svd autoencoder and generalised latent assimilation for nuclear reactor physics. Annals of Nuclear Energy **179**, 109431 (2022)
11. Hasegawa, K., Fukami, K., Murata, T., Fukagata, K.: Cnn-lstm based reduced order modeling of two-dimensional unsteady flows around a circular cylinder at different reynolds numbers. Fluid Dynamics Research **52**(6), 065501 (2020)
12. Ko, A., Lee, N., Sham, R., So, C., Kwok, S.: Intelligent wireless sensor network for wildfire detection. WIT Transactions on Ecology and the Environment **158**, 137–148 (2012)
13. Lever, J., Arcucci, R.: Sentimental wildfire: a social-physics machine learning model for wildfire nowcasting. Journal of Computational Social Science **5**(2), 1427–1465 (2022)
14. Lever, J., Arcucci, R., Cai, J.: Social data assimilation of human sensor networks for wildfires. In: Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments. pp. 455–462 (2022)
15. Loureiro, M.L., Alló, M., Coello, P.: Hot in twitter: Assessing the emotional impacts of wildfires with sentiment analysis. Ecological Economics **200**, 107502 (2022)
16. Mendon, S., Dutta, P., Behl, A., Lessmann, S.: A hybrid approach of machine learning and lexicons to sentiment analysis: Enhanced insights from twitter data of natural disasters. Information Systems Frontiers **23**, 1145–1168 (2021)
17. Pandey, A., Pokharel, R.: Machine learning based surrogate modeling approach for mapping crystal deformation in three dimensions. Scripta Materialia **193**, 1–5 (2021)
18. Qian, J., Lin, H.: A forest fire identification system based on weighted fusion algorithm. Forests **13**(8), 1301 (2022)
19. Smith, T.R., Moehlis, J., Holmes, P.: Low-dimensional modelling of turbulence using the proper orthogonal decomposition: a tutorial. Nonlinear Dynamics **41**, 275–307 (2005)
20. Staudemeyer, R.C., Morris, E.R.: Understanding lstm–a tutorial into long short-term memory recurrent neural networks. arXiv preprint arXiv:1909.09586 (2019)
21. Trunfio, G.A.: Predicting wildfire spreading through a hexagonal cellular automata model. In: Cellular Automata: 6th International Conference on Cellular Automata for Research and Industry, ACRI 2004, Amsterdam, The Netherlands, October 25-28, 2004. Proceedings 6. pp. 385–394. Springer (2004)
22. Wiewel, S., Becher, M., Thuerey, N.: Latent space physics: Towards learning the temporal evolution of fluid flow. In: Computer graphics forum. vol. 38, pp. 71–82. Wiley Online Library (2019)
23. Xu, R., Lin, H., Lu, K., Cao, L., Liu, Y.: A forest fire detection system based on ensemble learning. Forests **12**(2), 217 (2021)
24. Zhang, C., Cheng, S., Kasoar, M., Arcucci, R.: Reduced order digital twin and latent data assimilation for global wildfire prediction. EGUsphere pp. 1–24 (2022)
25. Zhu, Q., Li, F., Riley, W.J., Xu, L., Zhao, L., Yuan, K., Wu, H., Gong, J., Randerson, J.: Building a machine learning surrogate model for wildfire activities within a global earth system model. Geoscientific Model Development **15**(5), 1899–1911 (2022)
26. Zhuang, Y., Cheng, S., Kovalchuk, N., Simmons, M., Matar, O.K., Guo, Y.K., Arcucci, R.: Ensemble latent assimilation with deep learning surrogate model: application to drop interaction in a microfluidics device. Lab on a Chip **22**(17), 3187–3202 (2022)

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_33