Estimating Wildfire Duration using regression methods

Hansong Xiao
Imperial College London
`hx221@ic.ac.uk`

Abstract

Wildfire forecasting problems usually rely on complex grid-based mathematical models, mostly involving Computational fluid dynamics(CFD) and Celluar Automata, but these methods have always been computationally expensive and difficult to deliver a fast decision pattern. In this paper, we provide machine learning based approaches that solve the problem of high computational effort and time consumption. This paper predicts the burning duration of a known wildfire by RF(random forest), KNN, and XGBoost regression models and also image-based, like CNN and Encoder. Model inputs are based on the map of landscape features provided by satellites and the corresponding historical fire data in this area. This model is trained by happened fire data and landform feature maps and tested with the most recent real value in the same area. By processing the input differently to obtain the optimal outcome, the system is able to make fast and relatively accurate future predictions based on landscape images of known fires.

1 Introduction

As humans evolved, fire seems to become more controllable. But wildfires are not included, wildfires are a natural phenomenon that is difficult to predict and occur with high frequency, more than 4000 fires happened in CA, USA from 2003 to 2016 [1]. More seriously, wildfire can even be called a natural hazard [2]. Wildfires endanger most of the world[3], 4 million km^2 to 6 million km^2 (depends on different estimate methods) of land are affected by wildfire every year[4], which is almost half the area of Europe.

Wildfire has a huge negative impact on earth's ecosystem, risking natural systems and health, hydrogeomorphic effects from wildfire jeopardize the quality of daily water in both human society and the ecosystem[5]. Wildfire also threatens the world economy, house price of northwest Montana, a potential resort area in the USA, was depreciated by unpredictable long-term wildfire[6]. It also results in serious injury or death for animals, a huge group of sheep burned in a wildfire in 2006[7]. To prevent multiple hazards, researchers have begun to engage in developing wildfire prediction. Cellular Automata [8] and CFD (Computational Fluid Dynamic) [9] were widely used. These traditional methods are normally based on building a complex mathematical model that takes all variables and interactions between features into consideration. One of the most important and pioneering models is Rothermal's equation[10], but most of these models are hard to understand and lack of interpretability. Basically, a solution to the grid-based method is solving dynamic function or a system of partial differential equation[11], which is notoriously time-consuming. Thus, grid-based methods were intensive in computation and required resources that are nearly impossible to achieve a fast-decision pattern. With the development of artificial intelligence, machine learning becomes a new research direction[12–14] on wildfire prediction. Satellite imagery was used to help analyze geomorphic features to build models for better forecasting [15, 16]. However, it is challenging to develop a general and effective ML model in a worldwide range for forecasting wildfire, due to the complicity of environmental parameters in different areas.

Currently, most ML models put insight into a specific direction, like the occurrence of human-caused wildfire[17], forecasting sensitivity map of wildfires[18], and predicting the likelihood of wildfires [19]. Most of the current research is focusing on possible future wildfires, which requires management groups put energy in every possible location. 80 percent of fires last no more than one day based on

our extracted dataset [1], indicating fire management groups are unable to respond effectively in a limited time. It would be wise to keep eye on preventing potential large fires. In this paper, we aim to deliver a fast-decision model that focuses on predicting the likely extent of ongoing fires. ML models were split into two parts, simple regression models and image-based models. In the ML regression part, three methods were applied, including RF[20], KNN[21, 22], and XGBoost[23] to predict fire duration. In the image-based part, two CNN-based models were introduced, including multi-layer CNN and Encoder-Based CNN [24].

2 Data curation

2.1 Data Gathering

We train and validate our model using most of the fires that occurred in 13 years in a specific region, as shown in figure 1, ranging from 32.7N to 42N, -124.26W to -93.5W, which is mainly the Midwestern region of the United States, including California, Phoenix, Utah, Colorado, New Mexico, Kansas, and Texas. This part was chosen due to the similarity of geomorphology among them [25], the increasing number of wildfires with an expanded rate of a burned area at $355km^2$ per year [26], and an enormous economic loss excesses billions of dollars per year [27]. If we consider the entire U.S., the geomorphic disparity between the West Coast and the East and the sparsity of wildfire on the East would impact the accuracy of the model.

Figure 1: Target area

We decide to use *altas fire dataset*[1] for fire data, which includes the coordinates, start and end times, burned area, perimeter, and duration of the fires. We selected fire duration as the label value, considering the duration becomes more significant in evaluation wildfire[28]. Eighty percent of fires are small, and last less than one day. To avoid data imbalance, we randomly dropped half of the small fire (duration less than 1 day).

The database was split into two parts, from 2003 to 2013 as a train set, and from 2013 to 2016 as a test set. The ratio of fire number between the train and test is around 0.75, which is a reasonable number for data splitting. We expand each target coordinate in four directions for 30 kilometers range which is demonstrated in figure 2. The yellow point represents the coordinate of a random fire, and each red line shows how the 30 kilometers radius of the extension area. the square surrounded by green lines is the feature map to gather. We use a 30km radius size to avoid repeat data extraction work, which is time-consuming if a smaller radius were taken into consideration. Currently, three global scale databases were involved in this experiment, density map of above(or below) ground living biomass storage [29], landform slope map [30], land cover map of multiple surface characteristics (e.g. grass

coverage, tree coverage, snow converge) [31], and climate data corresponding to specific location range from 1979 to 2020 (e.g. precipitation, wind speed) [32]. In general, five input variables, tree cover, ${}^r\mathbf{X}_z^{tc} = \{{}^r x_{i,j}^{tc}\}$, geo-slope, ${}^r\mathbf{X}_z^{slope} = \{{}^r x_{i,j}^{slope}\}$, grass cover, ${}^r\mathbf{X}_z^{gc} = \{{}^r x_{i,j}^{gc}\}$, wind in u direction, ${}^r\mathbf{X}_z^{windu} = \{{}^r x_{i,j}^{windu}\}$ and wind in v direction ${}^r\mathbf{X}_z^{windv} = \{{}^r x_{i,j}^{windv}\}$ were obtained to predicting ${}^r\mathbf{D}^z$, duration. Here, i, j, the size of the matrix, for slope range from 0 to 200, grass cover and tree cover range from 0 to 700, z, and wind range from 0 to 4, fire index, r, the radius of extension. A detailed table of the dataset used is shown in table 1.

Dataset	Resolution	Sources	Image Size
Global Fire Atlas with Characteristics of Individual Fires	None	MODIS[1]	None
Global ALOS mTPI	270 Metres	ERGo[30]	(200, 200)
opernicus Global Land Cover Layers	100 Metres	GlobCover, LC-CCI, MODIS, PROBA-V [31]	(700, 700)
ERA5 Monthly Aggregates - Latest Climate Reanalysis	27830 Meters	ECMWF[32]	(4, 4)

Table 1: Dataset Detail

Figure 2: Example of fire extension

2.2 Data Standardization

Data standardization was applied for all inputs due to different scales of vegetation cover, normalizing all inputs in the same scale. *min-max* normalization [33]were adopted to resize all data into 0 to 1 range, as the equation is shown below, where X indicates each input matrix, where f is the feature name, r representing radius, and z regarding fire index.

$$ {}^r X_z^f = \frac{{}^r X_z^f - min({}^r X_z^f)}{max({}^r X_z^f) - min({}^r X_z^f)} \tag{1} $$

3 Methodology

This section discusses several models designed for fast-decision fire prediction based on different algorithms. To ensure consistency across the results, the input and output for models constructed on different bases are standardized. Three ML regression models are employed, maintaining the same input. Besides, the image-based ML model is used for comparison with the same input, input contains five possible different images, and five feature maps, which are grass cover, tree cover, landform slope, and two wind components(windu and windv). The output for these models is fire duration. We make use of three datasets, data from 2003 to 2013 in the target area regarded as train data, from 2013 to 2016 classified as test data.

The prediction result is hard to evaluate by Root Mean Square Error(RMSE) because the severity range of fire is not precise. In the numerical aspect, it would be a prediction with more than 50 percent error if a test data has a predicted value of 20 but with a label value of 30. However, in the aspect of fire prevention, both 20 and 30 indicate this fire exists for a long term, showing the seriousness of this fire, thus it could be defined as a reasonable prediction, which can be classified as the true prediction. To balance the ambiguous prediction result from numerical and prevention aspects, another measurement was introduced to evaluate all the possible outcomes, which is called an accuracy plot.

A prediction was regarded as a True Prediction(TP) if the prediction is greater than the critical value, otherwise, it is a False Prediction (FP). Calculating the ratio of positive out of total prediction, as the equation shown below equation 2, gives the accuracy of prediction. This ratio varies with different critical values, critical values with a range of 0 to the maximum label value. Thus, an accuracy plot was obtained and the lowest point in the accuracy curve presented the worst situation of each method.

$$Accuracy = \frac{TP}{TP + FP} \tag{2}$$

where TP is the true prediction, which has a prediction greater than the critical value, FP is false prediction, which has a prediction smaller than the critical value.

Meanwhile, another equation was used to compute RMSE of each single prediction, which gives the ratio between predicted days and the original fire duration.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i}^{n} (\frac{\left\| {}^{r}D_i^z - {}^{r}\hat{D}_i^z \right\|}{\left\| {}^{r}\hat{D}_i^z \right\|})} \tag{3}$$

where ${}^{r}D_i^z$ is the label value in radius r and fire index z, $\hat{y}$ is the prediction value.

3.1 Regression modeling

The Regression model aims to give a simple prediction. This model regards different standardized features (vegetation cover, slope of topographic map, wind, and precipitation)as input to predict fire severity. Output is a measurement variable that could represent the fire severity (fire size, fire duration), fire duration was used because it shows more accuracy rather than other label value. For the radius, 5km was adopted here as the figure 8 shows that radius has nearly no effect on the result.

3.1.1 Data preprocessing

We initialized the collected data and divided it into different ranges of inputs. In order to explore the affection of different ranges of input, we cropped the original image with a radius of 30km to a smaller radius, from the center point to the corresponding scale, and calculate the average value of each image in a different radius. Then, we obtained a series of average values with different radius. The average value is the input for the regression model. For example, the average value of 80×80 in the middle of the original map will be used as input if we want a range of 3km.

$$ {}^{n}X_f^r = \frac{\sum x_{ij}}{i \times j} \tag{4}$$

4

where X is the average value of the input, r is the radius, n is the fire index and f is the feature name. Below is a demonstration of the crop process when the radius is equal to three kilometers.

$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,600} \\ \vdots & \ddots & \vdots \\ x_{600,1} & \cdots & x_{600,600} \end{bmatrix} \xrightarrow{radius=3km} \begin{bmatrix} x_{270,270} & \cdots & x_{270,330} \\ \vdots & \ddots & \vdots \\ x_{330,270} & \cdots & x_{330,330} \end{bmatrix}$$

3.1.2 Random Forest regression model

We use random forest (RF) [20] as the first regression method. Random forest is an extension branch of the bagging algorithm, and RF is based on a series of Decision Trees(DTs). Random Forest develops models, which consist of a series of trees, that predict individually and separately. The output integrates the results of each predictor by a majority rule to forecast an output or one classification. In the regression study, different sets of samples were obtained from an initial dataset, the result is the average value of all the trees. The hyperparameters contain the number of individual predictors n, and the depth of each predictor d.

3.1.3 K-nearest neighbors algorithm model

K-nearest neighbors (kNN) is a supervised learning algorithm, working on both classifier and regression [21]. In classification, KNN assigns each single data point into the same category as the highest frequency of k nearest neighbors. In regression, calculating the average label value of k nearest point of the input train data to get the prediction value. In KNN model, input is the mean value of each feature map, and we keep using fire duration as output. The hyperparameter is the number of neighbors k.

3.1.4 XGBoost

Extreme Gradient Boosting [23] is another ensemble learning algorithm derived from Gradient Boost Decision Tree (GBDT) [34]. XGBoost takes every input data into account as each input results in one tree. XGBoost also considers the gradient of data, which saves calculation time. XGBoost gives regularized term for the loss function, which penalties the complexity of the model and compensates the learning term, thus XGBoost is more effective and more capable of preventing over-fitting. The researched hyperparameters are the number of estimator n, learning rate l, and max depth d.

3.2 Image Based Modelling

This part aims to explore the performance of the image-based model in predicting fire severity. To ensure the accuracy of forecasting, we tried permutation of different inputs and decide vegetation cover and landform slope are the final input as they give the best accuracy. The same output value or label value, fire duration, was selected to ensure the comparability between models. We make use of the image as input instead of the average value. All feature maps were resized to the same and placed in different channels of one data array as the traditional CNN. Moreover, the fire duration dataset from 2003 to 2013 was used as a train dataset and from 2013 to 2016 as a test dataset. A 5km data radius was selected to avoid unnecessary computation and keep the same radius as the regression ML model.

3.2.1 Data preprocessing

Different ranges of images are not interested in this model, since the radius presents little impact on final result. The data preprocessing method applied is resized, which scale all image to the same size by an interpolation method. Resize keeps the image the same but only change width and height. This approach calculates the value of a new image pixel based on the value of the original pixel points by bi-linear interpolation[35] shown below.

$$x_{ij} = [(1-p)x_{ij} + px_{i+1,j}](1-q) + [(1-p)x_{i,j+1} + px_{i+1,j+1}]q \tag{5}$$

where p, q is the fraction of distance between points, x is the pixel point, i, j is the index of the input point.

3.2.2 Multilayer Convolutional Neural Network (CNN)

CNN, one of the most famous machines learning neural network[36], extracted the significance feature from original data by mathematics operation convolution, which is used in this model for wildfire prediction. CNN has the ability to process large amounts of data due to the number of its neural, CNN could better recognize and learn for image type input. The structure of this model is shown in figure 3. ReLu and linear was used as activation function. A 3×3 kernel size and 32 channel numbers are selected in the convolutional layer to maximum keep the feature of the image. Two dropout layer is added in this model in case of over-fitting. To avoid gradient vanishing [37], we added a LeakyRelu layer after the dense layer. Detailed structures on this model are shown on table 2. For the hyperparameters, 100 epochs, 128 batch size is adopted and $adam$ optimizer is used.
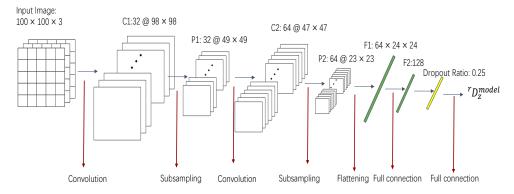

Figure 3: CNN Structure

Table 2: Structure of Multilayer CNN

Layer Type	Output Shape	Activation
Input	(100, 100, 5)	
Conv 2D(3×3)	(98, 98, 32)	ReLu
MaxPooling 2D(2×2)	(49, 49, 32)	
Conv 2D(3×3)	(47, 47, 64)	ReLu
Dropout(0.25)	(47, 47, 64)	
MaxPooling 2D(2×2)	(23, 23, 64)	
Flatten	36856	
Dense	128	linear
LeakyRelu	128	LeakyReLu(0.2)
Dropout(0.25)	128	
Dense	1	linear

3.2.3 Encoder Based Convolutional Neural Network (CNN)

Autoencoder [38] is an unsupervised machine learning technique to extract key features from input and ignore inessential points, learning the significant sets of data of input. Autoencoder consists of two parts, encoder and decoder, encoder responds to learning the crucial point from initial input, decoder is used to reduce refined features and return to its original state. In this model, we make use of an encoder, which is shown in table 3, to perform feature extraction. The encoder is trained jointly with an ML structure as shown in table 3. Two different encoders were applied in this model, due to the different sizes of the input image, vegetation map with a size of 100×100, and landform slope only has 30×30. Two encoders extracted features from the input separately and trained the concatenated dense layer. The detailed structure map is shown in figure 6 and table 3. In this study, the batch size, epochs, and optimizer were set to 128, 20, and $adam$, respectively.

6

Table 3: Structure of Encoder Based CNN

Layer Type	Output Shape	Activation
EncoderA		
Input	(100, 100, 1)	
Conv 2D	(100, 100, 32)	ReLu
MaxPooling 2D(2×2)	(50, 50, 32)	
Conv 2D	(50, 50, 16)	ReLu
MaxPooling 2D(2×2)	(25, 25, 16)	
Conv 2D	(25, 25, 16)	ReLu
MaxPooling 2D(2×2)	(13, 13, 16)	
Flatten	2704	
LeakyRelu	2704	LeakyReLu(0.3)
Dense	169	linear
LeakyRelu	169	LeakyReLu(0.2)
Dense	20	linear
EncoderB		
Input	(30, 30, 1)	
Conv 2D	(30, 30, 32)	ReLu
MaxPooling 2D(2×2)	(15, 15, 32)	
Conv 2D	(15, 15, 16)	ReLu
MaxPooling 2D(2×2)	(8, 8, 16)	
Flatten	1024	
LeakyRelu	1024	LeakyReLu(0.3)
Dense	64	linear
LeakyRelu	64	LeakyReLu(0.2)
Dense	10	linear
Encoder Based CNN		
Input A	(100, 100, 1)	
Input B	(30, 30, 1)	
Encoder A (Input A)	20	
Encoder B (Input B)	10	
Concatenate	50	
Dense	50	ReLu
Dense	50	Linear

3.2.4 Hyperparameter Tuning

To filter the optimal parameter value, grid search was applied to all models and compared the resulting loss on the test dataset. The number of estimators and max depth were studied for random forests. One more parameter, learning rate, was researched for XGBoost. Batch Size, epochs, and optimizer were tested for both image-based methods. Detailed results are shown in table 4.

4 Results and Discussion

The importance of radius was proved unnecessary as the Trend map shown in Appendix figure 8. Different Radius presents the similar impact on RMSE value of prediction, thus radius is not a key factor to consider when new models come. All mentioned ML methods were used to predict the fire duration on the test set, fire from 2013 to 2016 on the target area, producing the prediction map for the results from all regression-based ML approaches as shown in figure 4. Each plot has a predicted value on x label and a true value on y label and the scatter point was used to indicate individual forecasting. The blue scatter line shown on the map represents the real trend of fire duration. Additionally, figure 4(b) give the accuracy plots that provide another measurement to identify the preciseness of models. The feature importance of RF and XGBoost were shown in figure 7. The Outcome Summary can be seen in Table 5.

The result shown in Table 5 reveals that all applied ML methods presented no big difference on RMSE, although random forest performs the best. Moreover, all the models lack a reasonable R^2

Table 4: Optimal Parameter of Grid Search

Hyperparameters	Grid Search Space	Optimal Set
Random Forest		
Number of estimator	5, 10, 20, 50, 100	50
Max depth	1, 5, 10, 20, 50	5
KNN		
Number of neighbors k	5, 10, 20, 50	10
XGBoost		
Number of estimator	5, 10, 20, 50	50
Learning Rate	0.05, 0.1, 0.02	0.1
Max depth	1, 5, 10, 20, 50	5
Mutil-layer CNN		
Batch Size	32, 64, 128, 256	128
Epochs	10, 20, 50, 100	50
Optimizer	SGD, Adam	Adam
Encoder-based CNN		
Batch Size	32, 64, 128, 256	128
Epochs	10, 20, 50, 100	50
Optimizer	SGD, Adam	Adam

Table 5: Result Summary

Model	RMSE	$R^2 Score$	Accuracy
Random Forest	4.99	0.073	0.536
KNN	5.26	-0.031	0.572
XGBoost	5.05	0.049	0.535
Multi-layer CNN	5.05	0.051	0.629
encoder-based CNN	5.16	0.005	0.631

score for some reason, probably due to the large number of small fires, which incurs large data bias that is not well quantified by the R^2 score. Normally, a R^2 score above 0.3 presents a reliable "goodness-of-fit" of the model, but our results give R^2 score values all below 0.1, which indicates the inaccuracy of the models. However, some researchers argue that R^2 score may be inappropriate to evaluate the performance of a model and state that the accuracy of a model does not only depend on R^2 score[39]. Thus, the accuracy rate is more valuable than R^2 score for referencing the reliability of the model. In the ML regression model, KNN performs the best accuracy, achieving a worst situation around 0.572. RF and XGBoost have nearly the same accuracy rate due to the similarity of these two methods. For image-based models, below two figures 5(a) and 5(b) give the prediction and accuracy plot of two image-based ML models. It is clear that the image-based model demonstrates a more precise output, which improves by around 14 percent in accuracy. Although the two model presents similar result, they still have distinctiveness. Encoder-Based CNN is more suitable for scenarios in that different sizes of input are involved, multi-layer CNN is applicable to features of the same size. All plots performed an accuracy close to 90% at days equal to 10 and approaching 100% at days moving to 30, indicating all the models are more sensitive to long-lasting fire and more precise in predicting large fires.
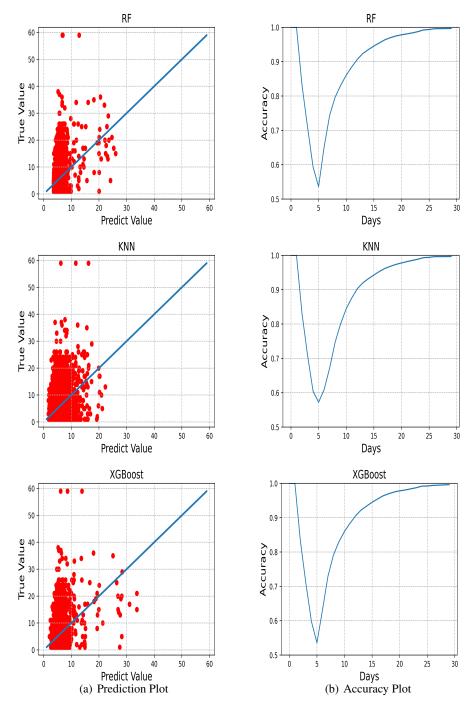
(a) Prediction Plot

(b) Accuracy Plot

Figure 4: regression model result

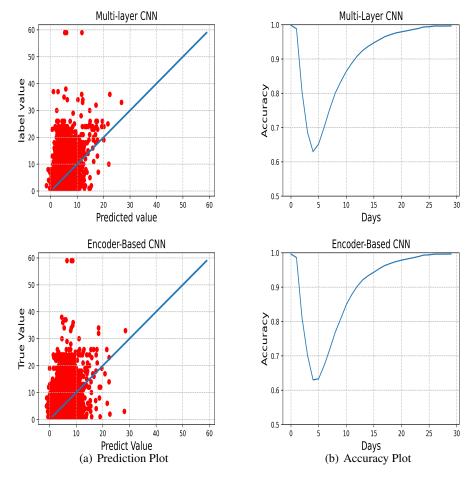(a) Prediction Plot   (b) Accuracy Plot

Figure 5: CNN model result

5 Conclusion

We developed a machine learning based wildfire prediction model that learns from historical fire data and corresponding landscape maps to achieve fast predictions. The average error of the model can be controlled within one day and has above eighty percent accuracy for fires that last longer than ten days. The models have difficulties in predicting extreme cases, such as the prediction of fires of very long or short duration. Besides, our model cannot achieve a good R^2 score as R^2 score cannot evaluate the result properly, another statistical coefficient is needed to replace R^2 score. This model is not suitable for the case that which small amounts of data and images were provided because our model requires amounts of input to learn a reasonable weight and bias. In addition, the input feature maps may change when applied to new target areas due to the feature importance of landscape features may change in a new region. Future enhancements to the model should focus on improving the regional coverage and generality and the simulation of features not considered in the current situation, utilizing various data augmentation techniques, and applying more encoders for different features.

Acknowledgments

References

[1] N. ANDELA, D.C. MORTON, L. GIGLIO, and J.T. RANDERSON. Global fire atlas with characteristics of individual fires, 2003-2016, 2019. 1, 2, 3

[2] Sarah McCaffrey. Thinking of wildfire as a natural hazard. *Society and Natural Resources*, 17 (6):509–516, 2004. 1

[3] William J. Bond and Jon E. Keeley. Fire as a global 'herbivore': the ecology and evolution of flammable ecosystems. *Trends in Ecology & Evolution*, 20(7):387–394, 2005. ISSN 0169-5347. doi: https://doi.org/10.1016/j.tree.2005.04.025. 1

[4] H. Youssouf, C. Liousse, L. Roblou, E.M. Assamoi, R.O. Salonen, C. Maesano, S. Banerjee, and I. Annesi-Maesano. Quantifying wildfires exposure for investigating health-related effects. *Atmospheric Environment*, 97:239–251, 2014. ISSN 1352-2310. doi: https://doi.org/10.1016/j. atmosenv.2014.07.041. 1

[5] François-Nicolas Robinne, Kevin D. Bladon, Carol Miller, Marc-André Parisien, Jérôme Mathieu, and Mike D. Flannigan. A spatial evaluation of global wildfire-water risks to human and natural systems. *Science of The Total Environment*, 610-611:1193–1206, 2018. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2017.08.112. 1

[6] Kyle M. Stetler, Tyron J. Venn, and David E. Calkin. The effects of wildfire and environmental amenities on property values in northwest montana, usa. *Ecological Economics*, 69(11):2233–2243, 2010. ISSN 0921-8009. doi: https://doi.org/10.1016/j.ecolecon.2010.06.009. Special Section - Payments for Ecosystem Services: From Local to Global. 1

[7] J. Madigan, J. Rowe, J. Angelos, W.F. Herthel, D. Matz, M. Dinucci, and V. Fletcher. (a323) wildfire associated burn injury of 1400 sheep in northern california: A coordinated mass casualty veterinary response. *Prehospital and Disaster Medicine*, 26(S1):s90–s91, 2011. doi: 10.1017/S1049023X11003074. 1

[8] Bastien Chopard and Michel Droz. Cellular automata. *Modelling of Physical*, 1998. 1

[9] John David Anderson and John Wendt. *Computational fluid dynamics*, volume 206. Springer, 1995. 1

[10] Richard C Rothermel. *A mathematical model for predicting fire spread in wildland fuels*, volume 115. Intermountain Forest & Range Experiment Station, Forest Service, US . . . , 1972. 1

[11] Alex Alexandridis, D Vakalis, Constantinos I Siettos, and George V Bafas. 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. 1

[12] Alexander L Fradkov. Early history of machine learning. *IFAC-PapersOnLine*, 53(2):1385–1390, 2020. 1

[13] Sibo Cheng, I Colin Prentice, Yuhan Huang, Yufang Jin, Yi-Ke Guo, and Rossella Arcucci. Data-driven surrogate model with latent data assimilation: Application to wildfire forecasting. *Journal of Computational Physics*, 464:111302, 2022.

[14] Sibo Cheng, Yufang Jin, Sandy P Harrison, César Quilodrán-Casas, Iain Colin Prentice, Yi-Ke Guo, and Rossella Arcucci. Parameter flexible wildfire prediction using machine learning techniques: Forward and inverse modelling. *Remote Sensing*, 14(13):3228, 2022. 1

[15] Caili Zhong, Sibo Cheng, Matthew Kasoar, and Rossella Arcucci. Reduced-order digital twin and latent data assimilation for global wildfire prediction. *Natural Hazards and Earth System Sciences*, 23(5):1755–1768, 2023. 1

[16] Sibo Cheng, Yike Guo, and Rossella Arcucci. A generative model for surrogates of spatial-temporal wildfire nowcasting. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2023. 1

[17] Marcos Rodrigues and Juan de la Riva. An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling & Software*, 57:192–201, 2014. ISSN 1364-8152. doi: https://doi.org/10.1016/j.envsoft.2014.03.003. 1

[18] Jake Lever, Sibo Cheng, and Rossella Arcucci. Social & physics based data driven methods for wildfire prediction. Technical report, Copernicus Meetings, 2023. 1

[19] Marc-André Parisien, Susan Snetsinger, Jonathan A Greenberg, Cara R Nelson, Tania Schoennagel, Solomon Z Dobrowski, and Max A Moritz. Spatial variability in wildfire probability across the western united states. *International Journal of Wildland Fire*, 21(4):313–327, 2012. 1

[20] Leo Breiman. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231, 2001. 2, 5

[21] N. S. Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992. doi: 10.1080/00031305.1992.10475879. 2, 5

[22] Helin Gong, Sibo Cheng, Zhang Chen, Qing Li, César Quilodrán-Casas, Dunhui Xiao, and Rossella Arcucci. 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. 2

[23] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016. 2, 5

[24] Sibo Cheng, Jianhua Chen, Charitos Anastasiou, Panagiota Angeli, Omar K Matar, Yi-Ke Guo, Christopher C Pain, and Rossella Arcucci. Generalised latent assimilation in heterogeneous reduced spaces with machine learning surrogate models. *Journal of Scientific Computing*, 94 (1):11, 2023. 2

[25] Tarek Kandakji, Thomas E. Gill, and Jeffrey A. Lee. Identifying and characterizing dust point sources in the southwestern united states using remote sensing and gis. *Geomorphology*, 353: 107019, 2020. ISSN 0169-555X. doi: https://doi.org/10.1016/j.geomorph.2019.107019. 2

[26] Philip E Dennison, Simon C Brewer, James D Arnold, and Max A Moritz. Large wildfire trends in the western united states, 1984–2011. *Geophysical Research Letters*, 41(8):2928–2933, 2014. 2

[27] Donald F Boesch, John C Field, and Donald Scavia. *The potential consequences of climate variability and change on coastal areas and marine resources: Report of the Coastal Areas and Marine Resources Sector Team, US National Assessment of the Potential Consequences of Climate Variability and Change, US Global Change Research Program*. Number 21. US Department of Commerce, National Oceanic and Atmospheric Administration . . . , 2000. 2

[28] David Radke, Anna Hessler, and Dan Ellsworth. Firecast: Leveraging deep learning to predict wildfire spread. In *IJCAI*, pages 4575–4581, 2019. 2

[29] Seth A. Spawn, Clare C. Sullivan, Tyler J. Lark, and Holly K. Gibbs. Harmonized global maps of above and belowground biomass carbon density in the year 2010. *Scientific Data*, 7(1):112, Apr 2020. ISSN 2052-4463. doi: 10.1038/s41597-020-0444-4. 2

[30] David M. Theobald, Dylan Harrison-Atlas, William B. Monahan, and Christine M. Albano. Ecologically-relevant maps of landforms and physiographic diversity for climate adaptation planning. *PLOS ONE*, 10:1–17, 12 2015. doi: 10.1371/journal.pone.0143619. 2, 3

[31] Marcel Buchhorn, Myroslava Lesiv, Nandin-Erdene Tsendbazar, Martin Herold, Luc Bertels, and Bruno Smets. Copernicus global land cover layers—collection 2. *Remote Sensing*, 12(6), 2020. ISSN 2072-4292. doi: 10.3390/rs12061044. 3

[32] H Hersbach, B Bell, P Berrisford, G Biavati, A Horányi, J Muñoz Sabater, J Nicolas, C Peubey, R Radu, I Rozum, et al. Era5 hourly data on single levels from 1979 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, 10, 2018. 3

[33] T Jayalakshmi and A Santhakumaran. Statistical normalization and back propagation for classification. *International Journal of Computer Theory and Engineering*, 3(1):1793–8201, 2011. 3

[34] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017. 5

[35] P.R. Smith. Bilinear interpolation of digital images. *Ultramicroscopy*, 6(2):201–204, 1981. ISSN 0304-3991. doi: https://doi.org/10.1016/0304-3991(81)90061-9. 5

[36] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)*, pages 1–6. Ieee, 2017. 6

[37] Stamatis Mastromichalakis. Alrelu: A different approach on leaky relu activation function to improve neural networks performance. *arXiv preprint arXiv:2012.07564*, 2020. 6

[38] Andrew Ng et al. Sparse autoencoder. *CS294A Lecture notes*, 72(2011):1–19, 2011. 6

[39] Charles Onyutha. From r-squared to coefficient of model accuracy for assessing" goodness-of-fits". *Geoscientific Model Development Discussions*, pages 1–25, 2020. 8

A Appendix

Main notations and acronyms

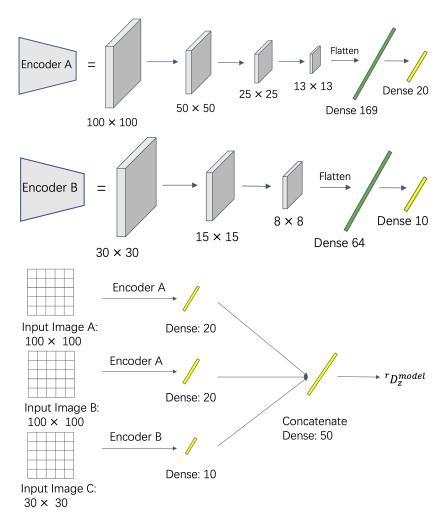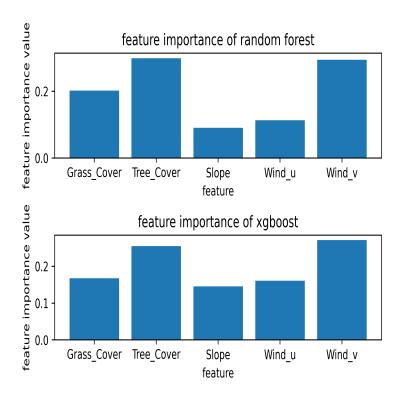|	
${}^{r}\mathbf{X}_{z}^{tc}$	Image of tree cover tc with fire index z in Radius r
${}^{r}\mathbf{X}_{z}^{gc}$	Image of grass cover gc with fire number z in Radius r
${}^{r}\mathbf{X}_{z}^{slope}$	Image of slope $slope$ with fire number z in Radius r
${}^{r}\mathbf{D}_{z}^{model}$	Predicted Duration of model with fire number z in Radius r
CNN	Convolution Neural Network
RF	Random Forest
KNN	K-Nearest Neighbors
$XGBoost$	eXtreme Gradient Boosting
CFD	Computational fluid dynamics

B Supplyment Figures

Figure 6: Encoder Based CNN Structure
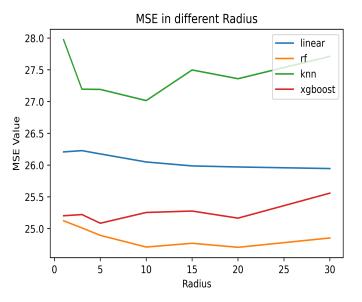
Figure 7: Feature Importance Plot

Figure 8: Trend map