

Wildfire risk assessment using remote sensing data

Giuseppe Marinelli

Graduate School of Natural Sciences (GSNS) - MSc Artificial Intelligence

Utrecht University

Utrecht, The Netherlands

Abstract—Assessing the risk of wildfires over the entire globe can be crucial in avoiding harm to wildlife, economy, properties and humans. This is known to be a challenging task. Here, a machine learning model is trained on a dataset composed of remote sensing data variables such as topography, vegetation and weather. The model is able to assess the risk of fire with a spatial resolution of 1000m/pixel. It achieves optimal results compared to other state-of-the-art architectures. Most of the variables in the dataset are found to be critical for the task, while few were disregarded. Particular focus has been given to collecting data across a variety of landscapes. Specifically, samples from Africa, Australia, Asia, Europe, South America and the US are included. This research shows the potential for deploying global wildfire risk assessment applications.

I. INTRODUCTION

Climate change is affecting the Earth with extreme events difficult to be foreseen. Wildfires are one of them. They are dangerous, devastating, occurring both naturally or human-induced. These events damage the economy, properties, wildlife and environment. Mizutori and Guha-Sapir (2017) mention that from 1998 to 2017 there were 254 wildfire disasters around the world. The economical loss for these events is estimated at 68 billion US dollars (nearly the total GDP of Luxembourg in 2017). The First Street Foundation (2022) published a report for the year 2022 stating that in the US, nearly 72 million homes have an average risk of being damaged by a wildfire, while 4 million properties have an extreme or severe risk. Chen et al. (2021) identified that short-term exposure to wildfire-related fine particulate matter (PM2.5) in the air is linked to an increased risk of mortality. Climate change also contributes to forest decline, increasing water scarcity, water evaporation and tree mortality (Park Williams et al. 2013). All these factors contribute to increasing the available fuel source for wildfires.

The National Interagency Fire Center (NIFC) (2021) of the United States, reported 58,950 wildfires across the country. For a total of $40,963 \text{ Km}^2$ burned (nearly the size of the Netherlands). Wildfire records reveal that in the last 10 years the number of wildfires in the US has decreased. However, the area of impact has increased, causing more damage. A report additionally showed that the extent of area burned by wildfires each year seems to have increased since the 1980s (US EPA 2021).

This thesis was prepared in partial fulfilment of the requirements for the Master Degree of Science in Artificial Intelligence, Utrecht University. Supervisor(s): Mathieu Gravey and Marc J. van Kreveld.

Given the rising damage of wildfires and an ongoing climate change process, prevention is the only means of managing and reducing costs to humans and nature. Stein et al. (2013), state that communities with strong wildfire prevention programs are likely to have fewer human-caused ignitions.

It is more expensive to fight fires and rebuild what was destroyed than it is to avoid them. This can be done by educating the community, by instructing the government to provide funds, equipment and expertise, essential for preparedness in case of emergency. More can be done also by monitoring the condition of trees, cutting trees near lines, burning critical fuel, monitoring the weather and soil condition (Stein et al. 2013). Given many factors that contribute to the start of a fire, such as human intervention, wind, soil and air moisture, predicting where a wildfire could start is challenging. Forecasting the risk or the area in which a wildfire can occur could lower the impact on the economy and ecologic damage (Jazebi, De Leon, and Nelson 2019, Surya 2020).

Predicting the path, intensity or occurrences of wildfires is difficult because of complicated meteorological scenarios, complex terrain effects on the airflow and spatially heterogeneous and physically elaborated fuel structures. Modelling and predicting wildfires is a multi- and inter-disciplinary challenge addressed by various fields such as engineering, ecology, physics, atmospheric science, chemistry, mathematics, forestry, and other fields, all linked by computational science (Coen and Douglas 2010).

Historical fire data is hard to rely on due to various biases, such as limited observations or reporting. It also requires hand-labelling, a time consuming task. Remote sensing is the process of detecting and monitoring some physical characteristics of an area by measuring its reflected and emitted radiation at a distance, typically from satellites (*What is remote sensing?* U.S. Geological Survey (USGS) (2021)). It provides reliable and frequent data that overcomes the historical data's limitations. A copious amount of data allows researchers to build more robust models that can be applied on a large scale. However, building a model relying purely on satellite data is sub-optimal. Satellites cannot cover the entire earth in one single instance but need to rotate around the globe. Their data can also be obstructed by clouds. These issues can be partially solved by estimating the data from ground data or inferred from past data.

In recent years, studies constructed models and gathered different types of remotely sensed products in order to predict wildfires.

The first quantitative tool for predicting the spread and intensity of forest fires was developed by Rothermel (1972). The model attempted to mathematically describe the physical and chemical processes of fire. It required variables such as fuel, winds, slopes, moisture, climate and weather. Although this model was not complex, it was reliable and easy to use. This was an essential component to predict the fire spread and intensity on the spot (Wells (2008)).

The Global Fire Atlas presented in the paper by Andela et al. (2019), is a worldwide dataset that allows to dynamically track individual fires, tracing their size, duration, daily expansion, fire line length, speed, and direction of spread. The data used in the research is derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) Collection 6 burned-area dataset. Such algorithms are useful for fire management and investigation of the vegetation–fire feedback, as well as climatic and human controls on global burned area.

Artés et al. (2019), developed a data mining algorithm (GlobFire) to create a global wildfire database to pursue analysis on fire regimes and fire behaviour. The input to the model is composed of the burnt area Collection 6 (MCD64A1) from the NASA's MODIS. GlobFire characterizes the fires and their perimeters. It is comparable to the Global Fire Atlas, but it focuses more on the fire patch behaviour than on the characterization of the fire events. Although their goals are similar, they rely on a different methodology. The algorithm is available on the Global Wildfire Information System platform.

With recent progress of Artificial Intelligence (AI) in the field of Machine Learning (ML), the task of monitoring and predicting a wildfire can benefit from such advances. ML can identify highly complex patterns ignored by other methods. ML techniques require a vast amount of data to be trained. These algorithms are commonly trained using supervised or unsupervised learning. The first requires a label linked to each sample in the dataset. Labelling is usually done manually to ensure that the model will be properly trained. The second method does not require labels. The model learns to group together samples by their properties, inferring knowledge from these.

Sayad, Mousannif, and Al Moatassime (2019), built a model to predict wildfire occurrences in Canada between 2013 and 2014. The dataset is composed of data relative to the state of the crops, the state of the soil, and a fire indicator. The last is taken from the *MOD14A1 V6 dataset*. It provides a daily fire mask composites at 1km resolution derived from the MODIS 4- and 11-micrometer radiances. It gives direct information about the confidence of fire detected (Justice et al. (2002)). They collected 804 instances (386 positive and 418 negative labels). After preprocessing the dataset, the team trained a Multi-Layer Perceptron (MLP) Classifier and a Support Vector Machine (SVM) Classifier algorithm. Both models achieved a high score on accuracy and F1-score (respectively 98.3% and 97%). Their results are promising for further research on wildfire prediction using Neural Networks (NNs).

A second approach that applies machine learning was proposed by Radke, Hessler, and Ellsworth (2019). The team

leveraged deep learning to predict wildfire spread. The algorithm called “FireCast” is composed of a 2D Convolutional Neural Network (CNN) that is trained on supervised data. FireCast merges AI and Geographic Information Systems (GIS) with satellite imagery, elevation data, weather data, and historical fire perimeters. The model predicts the area that is going to burn during the following 24 hours given the initial inputs. Such algorithms reduce the computation time required for critical decisions. Methods such as the Wildland Fire Decision Support System (WFDSS)¹ requires a large amount of appropriate data. The team had to interpolate weather data and apply data augmentation to generate a sufficient amount of training data. FireCast can outperform traditional wildfire modelling software, with an average accuracy of 87.7%.

Following the previous approach, Huot et al. (2021) constructed a dataset containing historical wildfire records, topography of the US, weather data (surface temperature, precipitation, winds, and humidity), USA drought data, vegetation indices, population density and an energy release component. Their dataset contains 18,545 samples, 58% in which the fire increases in size, 39% in which the fire decreases in size and for the remaining 3%, the fire stays the same size. Given the element of time introduced as a variable in the dataset, the team used an autoencoder, as it would be more effective than a simple CNN. This type of machine learning algorithm is a sequence to sequence model. It is commonly used for tasks such as machine translation, voice recognition and video captioning. An autoencoder can map sequences of different lengths to each other. Through this means, the team can achieve good results in predicting the *next day wildfire* (Huot et al. 2021). The model was more accurate when it was given lower resolution images than when it was given higher resolution ones. It was able to detect the correct next day fire pixels with an accuracy of 35% for images 1 km x 1 km and 67% for images 8 km x 8 km. The team considers a fire pixel as a fire, regardless of its intensity.

Overall, previous research showed that remote sensing data is an essential tool to predict wildfires globally. This data can be obtained daily from satellites, as well as being estimated. This work builds on the research of Sayad et al. by:

- designing a ML architecture with the prospect of near real-time capabilities;
- comparing the proposed model with other state of the art models;
- training a model capable of assessing the risk of wildfire across a variety of landscapes;
- investigating a range of remote-sensing data sources from GEE;

The first step toward dealing with such a complex task is constructing a dataset. This is composed by images extracted from Google Earth Engine (GEE). After gathering and preprocessing the dataset, it is split into training and validation. The dataset is fed into a CNN that classifies each image's pixel as

¹The WFDSS is a web-based geospatial fire management portal used by some state and federal fire agencies to manage and document large fires



Figure 1: The image shows the location of the samples that compose the training and validation sets.

1 (probability of fire in the area) or 0 (no probability of fire in the area). In the next sections an explanation of the features in the dataset will follow, section II. The ML architectures used are explained in section III. After describing the experiments in section IV, the results are presented in section V and then discussed in section VI. Finally, conclusions are drawn in section VII based on the data gathered in the experiments.

II. DATA

Using GEE, each image is extracted from different areas around the globe, from the years 2018, 2019 and 2020. In particular, focus has been given to fire seasons in Africa, Asia, Australia, Europe, South America and USA (an example is shown in Figure 1).

The features extracted from GEE have been chosen with a broad geographical and historical coverage, as well as with regards to missing data. Previous research composed their dataset with only two features. Here, 20 remote-sensing data sources are investigated. Spatial and temporal resolution details about the features are shown in Table I.

- **Elevation** data is from the NASADEM dataset, a modernized collection of the Digital Elevation Model (DEM) and associated products generated from the Shuttle Radar Topography Mission (SRTM) data (Crippen et al. 2016). The elevation of water bodies is not included in this dataset.
- **History LAI and FAPAR** data are from the NOAA Climate Data Record. Its values are computed globally over land surfaces, but not over sparsely vegetated areas, permanent ice or snow, permanent wetland, urban areas, or water bodies. The history of the data refers to a 10 year mean of Leaf Area Index (LAI) and Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) in the month in which the image has been extracted. This feature holds information about the vegetation status of the area.
- **Land Surface Temperature (LST)** data is from the ERA5-Land dataset. It provides an evolutionary view of

Image feature	Spatial resolution	Temporal resolution
Elevation	30 m/pixel	1 year
History LAI	5,566 m/pixel	10 years
History FAPAR	5,566 m/pixel	10 years
LST	11,132 m/pixel	daily
History LST	4,638 m/pixel	5 days
Soil temperature	11,132 m/pixel	daily
History soil temperature	11,132 m/pixel	5 days
Daily precipitations	5,566 m/pixel	daily
History precipitations	5,566 m/pixel	5 days
Air pressure	11,132 m/pixel	daily
Wind u component	11,132 m/pixel	daily
Wind v component	11,132 m/pixel	daily
Daily humidity	11,132 m/pixel	daily
History humidity	11,132 m/pixel	5 days
Daily LAI high	11,132 m/pixel	daily
Daily LAI low	11,132 m/pixel	daily
Daily NDVI	463 m/pixel	daily
8 days Evapotranspiration	500 m/pixel	8 days
History fire	1,000 m/pixel	1 year
Land cover	500 m/pixel	daily

Table I: The table lists all 20 image features included in the dataset, with their respective spatial and temporal resolutions.

land variables over several decades. ERA5-Land data is available from 1981 up to three months from real-time. Land surface temperature is measured at 2m above the surface. The maximum daily temperature is obtained for this feature.

- **History LST** data is from the Global Change Observation Mission (GCOM) dataset. This feature consists of the mean land surface temperature over five days prior to the image being captured.
- **Soil temperature and history soil temperature** data are from the ERA5-Land dataset. This is the temperature of the soil in layer 1 (0 - 7cm). The daily maximum soil temperature is extracted for every image. The history is then the mean soil temperature value of five days.
- **Daily precipitations and history precipitations** data are from the Climate Hazards Group InfraRed Precipitation with Station dataset (CHIRPS). It is a 30+ year quasi-

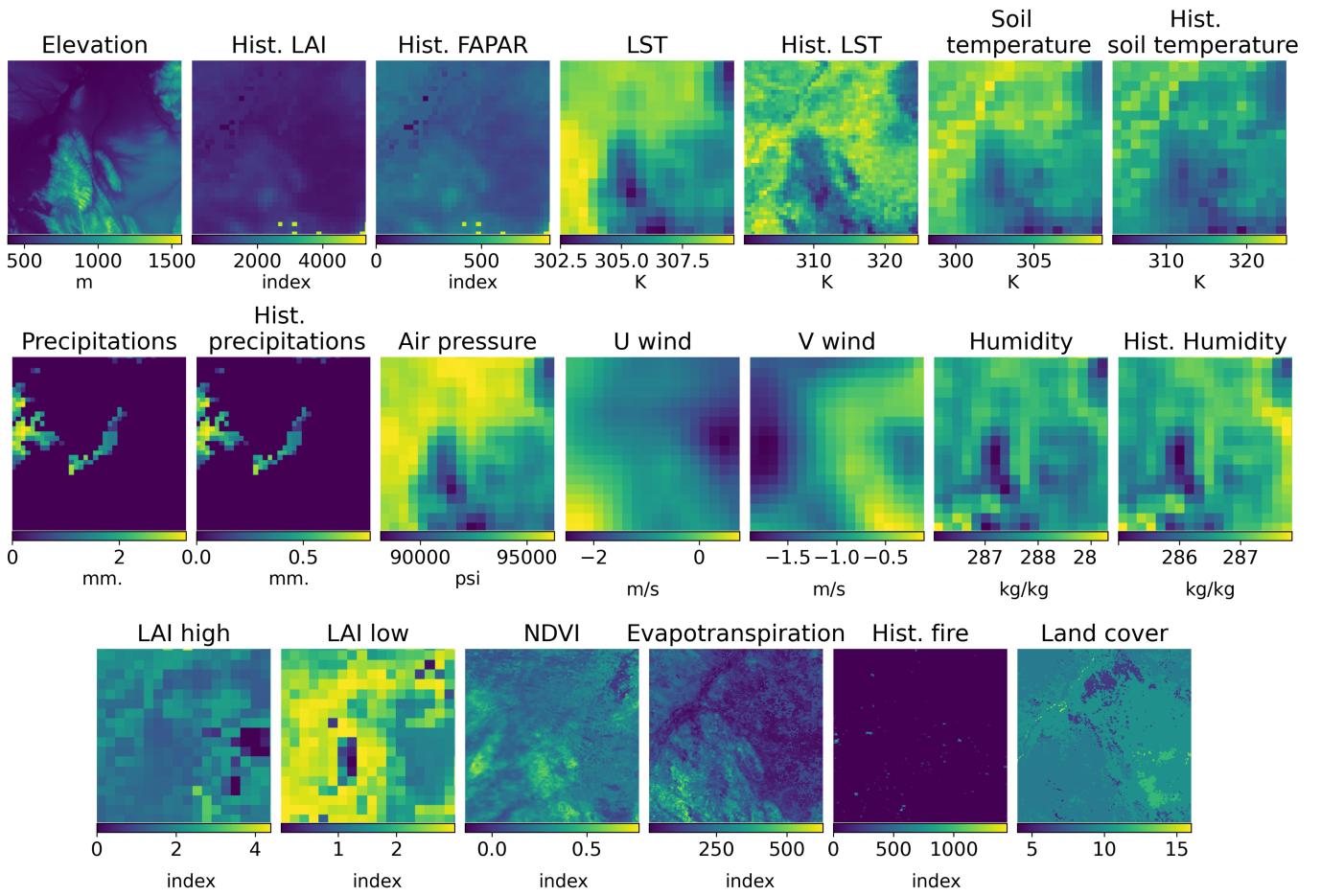


Figure 2: The graph shows a sample from the training set of the 20 remote-sensing data sources extrapolated from the GEE. Each image is 200 by 200 pixels.

global rainfall dataset. The precipitation history is the mean value of rainfall over five days.

- **Air Pressure** data is from the ERA5-Land dataset. The daily minimum air pressure is obtained for every image.
- **Wind u and v component** data is from the ERA5-Land dataset. Both components give information about the wind direction. The daily mean of the two components is extracted.
- **Daily humidity and historical humidity** data are from the ERA5-Land dataset. A daily mean of the first components is extrapolated, while the history is a five days mean.
- **Daily LAI high and daily LAI low** data are from the ERA5-Land dataset. LAI low and high stands for respectively high and low vegetation type. It is obtained a daily minimum value of the two components.
- **Daily NDVI** data is from the MODIS dataset. The Normalized Difference Vegetation Index (NDVI) quantifies vegetation by measuring the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs) (*What is NDVI* 2022).
- **8 days Evapotranspiration** data is from the MODIS dataset. Evapotranspiration/Latent Heat Flux product is an 8-day composite product. It is the sum of evaporation

from the land surface plus transpiration from plants. The pixel value for the Evapotranspiration is the sum of all eight previous days (*Evapotranspiration and the Water Cycle* 2018).

- **Fire history** data is from the MODIS Terra Thermal Anomalies & Fire Daily Global 1km dataset. This is the maximum intensity per-pixel of fires which happened over a period of 1 year prior to the image. An area devastated by a fire is unlikely to be lightened up again, since it is lacking fuel to burn.
- **Land cover** data is from the MODIS Land Cover Type Version 6 dataset. This is derived using supervised classifications of MODIS Terra and Aqua reflectance data. The band *Land Cover Type 1* is extracted to differentiate between flammable and not flammable land (such as water bodies, urban areas, wetlands, barren and permanent snow or ice).

For each image, its label showing the pixel-fire instances is retrieved from the MODIS Terra Thermal Anomalies dataset. A fire is converted into a fire pixel regardless of its intensity. The dataset is naturally highly unbalanced, as the dataset presents more no-fire instances than fire instances (Table VII). Before feeding the dataset into the model, all images are *preprocessed*. In GEE clouds are exported as empty pixels.

On average, all features present 4% of missing pixels. Empty pixels replaced with zeros or the mean, depending on the feature:

- The *elevation* feature presents empty pixels corresponding to water bodies or irregularities. These are replaced with zeros.
- Empty pixels in the *history fire* feature are replaced by zeros because of the uncertainty of being a fire.
- For the *land cover* feature, empty pixels are treated as a non flammable area.
- All missing values from other features are replaced with their mean. This is due to constraints on time and easiness of implementation.

The dataset is composed of 900 instances for the training set and 150 for the validation set. Each image is 200 x 200 pixels, with a resolution of 1,000 m/pixel. The area distribution of the datasets is shown in Table II.

Dataset	Africa	Australia	Asia	Europe	South America	USA
Training set	491	137	0	0	196	76
Validation set	0	0	0	16	0	134

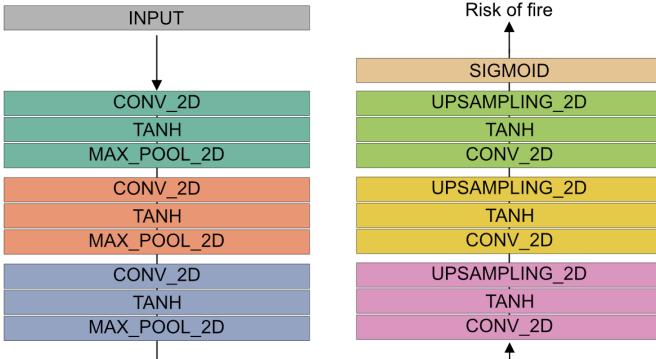
Table II: Geographical area distribution in the datasets.

The distribution depicted in Table II is not equally split due to the possibility of finding the same fire instance in multiple samples.

III. METHODS

Sayad et al. built a NN and a SVM, achieving over 97% accuracy in predicting wildfires in Quebec. This paper studies the risk of wildfire as an image segmentation task. Instead of classifying the entire picture, the model determines if each pixel is at risk of fire. The size of the model's output is the same as its input. Therefore, the prediction is an image mask, marking fires over the area of interest. This section explains how Autoencoder, U-Net and CNN architectures are configured, which loss functions are used, and which classification metrics are implemented.

A. Autoencoder architecture



same number of filters. The last layer converts the outputs of the preceding ones to a single filter through a *sigmoid* activation function. This U-Net architecture is borrowed from the library Tensorflow (Martín Abadi et al. 2015). The model has a total of 33,674,753 parameters.

C. CNN architecture

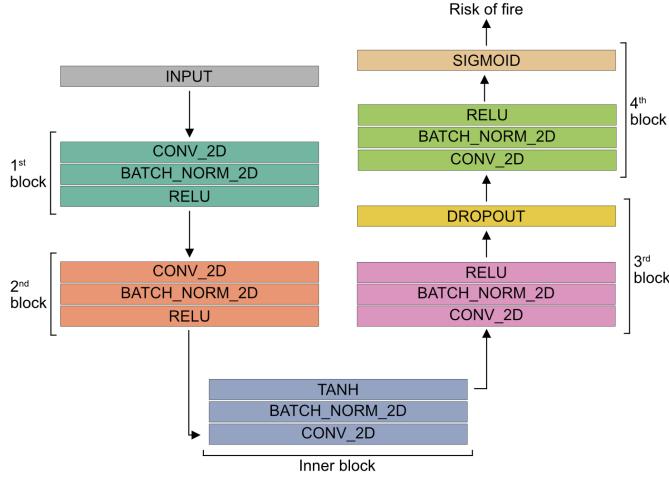


Figure 5: Architecture of the CNN.

The CNN architecture consists in multiple blocks of convolutional, batch normalisation, and activation layers as illustrated in Figure 5. Each convolutional layer retains the original size of the image while increasing the number of filters. These are 40, 60, 30 and 1 respectively for the first, second, third and fourth block. The inner blocks uses a *tanh* activation function while the other blocks rely on the *ReLU* activation function. The number of inner blocks is optimised in section IV. The last layer converts the outputs of the precedent ones to a single filter through a *sigmoid* activation function. The model has 587,177 parameters in total.

D. Loss function

Multiple loss functions such as Mean Squared Error (MSE), Binary Cross Entropy (BCE) and Focal Loss were tested. These showed poor results and therefore do not seem to be suited for this specific tasks. On the other hand, the *dice loss*, a loss function adapted from the dice coefficient, showed optimal results. This has been implemented along the three architectures presented previously. The dice coefficient is mostly used as a metric in computer vision to calculate similarities between two images (Jadon 2020). Dice score measures the relative overlap between the prediction and the ground truth. Its value is independent of the size of objects, i.e. large objects contribute to its value as much as small ones.

Given y_{pred} as the image predicted from the model and y_{true} as the label, the dice loss is calculated by first computing its coefficient as follows:

$$DICE_COEF(y_{pred}, y_{true}) = \frac{2 \cdot |y_{true} \cdot y_{pred}|}{|y_{true}| + |y_{pred}|} \quad (1)$$

$$DICE_LOSS = 1 - DICE_COEF \quad (2)$$

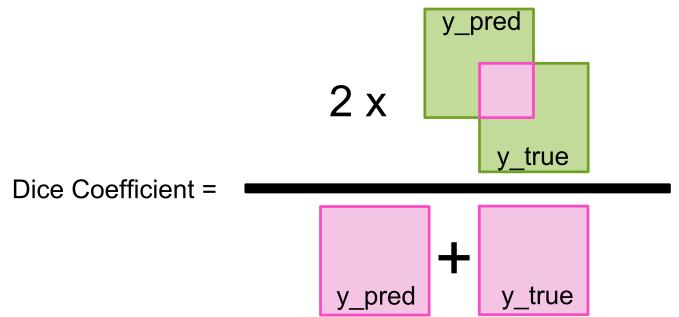


Figure 6: Graphic representation of the dice coefficient.

E. Classification Metrics

Metrics such as accuracy, precision and f1 score metrics are not suited for this task. Instead, four classification metrics have been used during the training and evaluation phases. The first, *fire ratio*, calculates the ratio of correctly predicted fires instances over the total number of true ground fires instances. In binary classification, this metric is commonly called *sensitivity*. Before calculating the *fire ratio*, the output of the model undergoes a threshold (set to 0.5) to translate the predictions to a binary value:

$$thr(image) = \begin{cases} 1 & \text{if pixel} > 0.5 \\ 0 & \text{else} \end{cases} \quad (3)$$

$$\text{FIRE_RATIO}(y_{pred}, y_{true}) = \frac{|thr(y_{pred}) \cdot thr(y_{true})|}{|thr(y_{true})|} \quad (4)$$

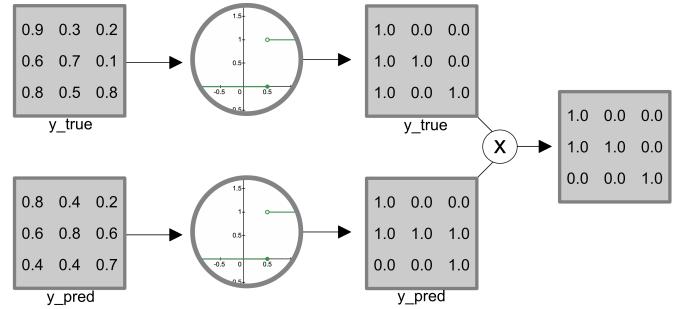


Figure 7: The figure shows an example of the process to calculate the fire ratio of a predicted image, given its ground truth. The fire ratio of this example is 0.8.

The second is the Intersection Over Union (IOU). It also known as the Jaccard similarity coefficient (Rahman and Wang 2016). It is used to describe the extent of overlap of two sets. For instance, it can be found in object detection tasks where a model is trained to fix a box around an object. A visual representation of the IOU formula is shown in Figure 8.

IOU is implemented as follows:

$$\text{IOU}(y_{\text{pred}}, y_{\text{true}}) = \frac{|abs(y_{\text{pred}} \cdot y_{\text{true}})|}{|y_{\text{true}}| + |y_{\text{pred}}| - |abs(y_{\text{pred}} \cdot y_{\text{true}})|} \quad (5)$$

with abs as the absolute value.

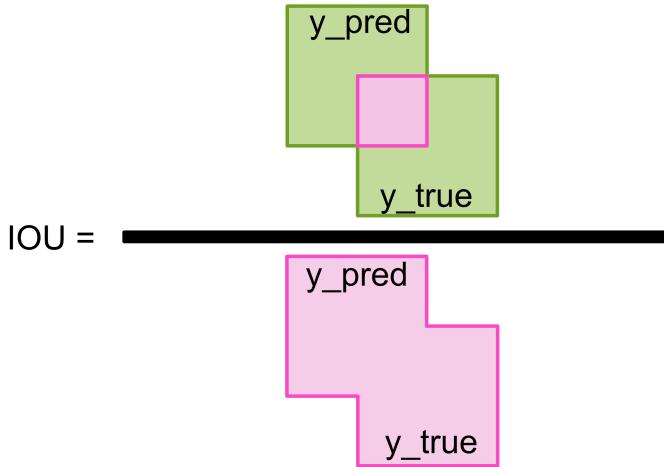


Figure 8: Graphic representation of the IOU.

The formula for the fourth metric, the *no fire ratio*, first converts the y_{pred} and y_{true} to their opposite (Eq.6), then it follows a similar formula as the *fire ratio*.

$$image_I(image) = \begin{cases} \text{if pixel} > 0.5 & 0 \\ \text{else} & 1 \end{cases} \quad (6)$$

$$\text{NO_FIRE_RATIO}(y_{\text{pred}}, y_{\text{true}}) = \frac{|thr(y_{\text{pred}}_I) \cdot thr(y_{\text{true}}_I)|}{|thr(y_{\text{true}}_I)|} \quad (7)$$

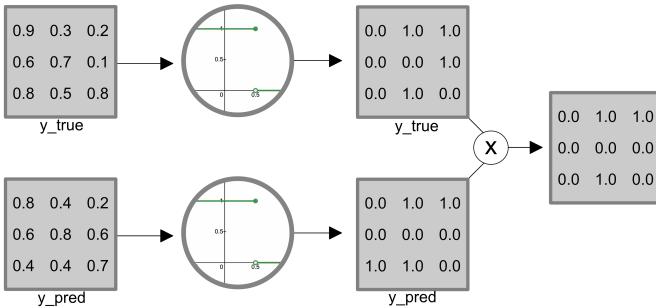


Figure 9: The figure shows an example of the process to calculate the no fire ratio of a predicted image, given its ground truth. The no fire ratio of this example is 0.75.

The fifth metric is the *dice coefficient*, Eq.1. The last two metrics are rough indicators for the model's training progress.

The *no fire ratio* is used to validate the model in the case of overfitting or underfitting. For instance, a high value of *no fire ratio* and a low value for the *fire ratio* shows that the model is reaching a local minimum by predicting only zeros (the most present class in the dataset). Training the model to achieve a high *no fire ratio* results in an unrealistic model. Given the dataset distribution, independent factors that could cause a fire (i.e. human intervention, sudden weather changes, climate change) or missing data, the possibility that a wildfire could start in a *non fire* pixel area is realistically always higher than zero (excluding areas in which a fire cannot ignite). Therefore, training the model only on the assumption that certain conditions result in a non fire area could be biased and realistically not possible.

IV. EXPERIMENTS

A. CNN Optimisation

To achieve optimal results for the CNN architecture, as seen in Figure 5, parameters such as number of inner blocks, number of incremental filters in the inner blocks' convolutional layers, learning rate and dropout are optimised following a grid search method. Everything is programmed in Python using tensorflow (Martín Abadi et al. 2015) on Colab with a single Tesla T4. The following values are explored:

Parameter	Values
No. inner blocks	{0, 2, 4, 6, 8}
No. incremental filters	{10, 20}
Dropout	{0, 0.1, 0.3, 0.5}
Learning rate	{0.01, 0.001, 0.0001}

Table III: Search space for the CNN's parameters.

By incremental filters is meant that the number of filters per successive convolutional layer increases by an x amount. Each combination of parameter runs for 30 epochs. The run time is between 15 min to 1 hour. Runs are pruned if the *fire ratio* of the epoch is 0 after 4 epochs. This behaviour is due to the model getting caught in a local minima. After ruling out bad parameter combinations, the optimal models are run for 5 times, creating a confidence measurement of their results.

B. CNN variability

Every time a CNN is initialised, new weights are assigned to its inner layers. This creates variability in the predictions. To validate the model, the variability of the CNN is calculated by initialising and training 10 different models with the optimised parameters defined in section IV-A. The computation of the results follows these steps: first, the output of the predictions undergo a threshold as shown in Figure 9. Secondly, the output of 10 models is summed for each image. Thirdly, The mean of each picture is calculated. An illustration of this process is shown in the following Figure 10.

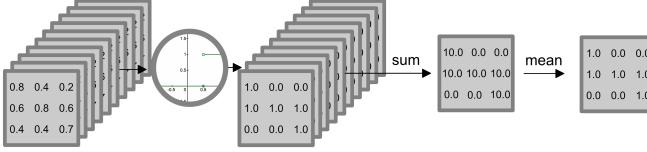


Figure 10: The picture shows an example of how the CNN variability is calculated given 10 outputs from the different models.

C. Comparing architectures

In this section, the CNN architecture is compared to a U-net and an autoencoder. The latter two models are trained with the same dataset as the CNN. U-net and autoencoder are optimised with a grid search approach. The search space for the parameters is listed in Table IV. The two optimiser used for the two models are the Adaptive moment estimation (Adam) and the Stochastic gradient descent (SGD).

Parameters		
Model	Optimisator	Learning rate
Autoencoder	{Adam, SGD}	{0.01, 0.001, 0.0001}
U-Net	{Adam, SGD}	{0.01, 0.001, 0.0001}

Table IV: Search space parameters for the autoencoder and the U-Net.

D. Feature Analysis

In this paragraph, the model's response to input adulteration is analysed. The aim of this experiment is to reduce the number of redundant features in the dataset and understand which of these are relevant to predict a risk of wildfire. Decreasing the size of the dataset also allows to increase the total number of trained samples. Machine learning methods are known for being data-eager. Therefore, more samples could increase accuracy and generalisation of the model. This experiment is performed on the evaluation set. For each of the 19 features and for each 10 variation values, a new evaluation dataset is created. Given the validation set, for each feature a minimum and a maximum value is extracted by examining all samples. These are called $feat_{min}$ and $feat_{max}$. A positive or negative amount of variation to the feature is calculated as follows:

$$\text{variation(amount)} = \frac{feat_{max} - feat_{min}}{100} * amount \quad (8)$$

adulterated_feature =

$$\begin{cases} \text{if } feature + \text{variation(amount)} > feat_{max} \quad feat_{max} \\ \text{if } feature + \text{variation(amount)} < feat_{min} \quad feat_{min} \end{cases} \quad (9)$$

After calculating the new value of the feature, this is constrained between $feat_{min}$ and $feat_{max}$.

The features *land cover* is excluded from this experiment because it would require more investigation and does not suit the format of this experiment.

After the feature analysis, the CNN is trained without features that are assumed to be redundant. This step validates the previous analysis.

E. Spatial dependence

Creating a wildfire risk assessment model able to work over the entire planet is challenging. The dataset fed to the model needs to include a variety of samples from different areas of the world. Multiple factors could ignite a fire in different parts of the world: human intervention, lack of fire management, weather, etc. In this section, the area's characteristics influence on the model's ability to generalise and predict the risk of a wildfire is studied. Asia is added to this test. Each set of tests is done by first excluding one geographical area (this composes the validation set, 150 samples), then an equal number of samples per area constructs the training set (900 samples). Figure 18 depicts the location for each sample in the dataset of this experiment.

V. RESULTS

All results presented in this section are computed on the validation set. The latter has been constructed with regard to samples overlap, being independent from the training set.

Figure 11 shows five areas in which the model predicts wildfires. Starting from the first image to the left. The CNN is able to locate the area near to the fire, although covering it partially. The second picture covers one of the fires entirely while missing the other. For the third image, the model catches all the fire instances overshooting. In the fourth sample the model is not able to locate the fire. In the last picture, the model successfully cover the location of the fires.

A. CNN optimisation

The CNN architecture has been designed with the prospect of near real-time capabilities. Following a grid search method with the parameters in Table III, the model is optimised. For practical reasons, the results in Table V are split between one of the number of filters.

Models	Fire ratio	No fire ratio	Dice coefficient	IOU
Model 1	0.82	0.87	0.0059	0.0029
Model 2	0.80	0.86	0.0059	0.0029

Table V: The table presents results for the CNN parameters' optimisation. For practical reasons, runs were split between 10 and 20 filters. Model 1 has parameters: 6 blocks, 10 incremental filters and a dropout of 0.3. Model 2 has parameters: 4 blocks, 20 incremental filters and a dropout of 0.5.

Initial results show that the models with 0 inner blocks and a dropout of 0 underfit. These are excluded from deeper investigations. After 20 epochs, all models highly fluctuate with all three learning rates, but optimal results are achieved with the two lowest learning rates. Therefore, an exponential decaying learning rate with an initial value of 0.0008 is set for all optimisation runs.

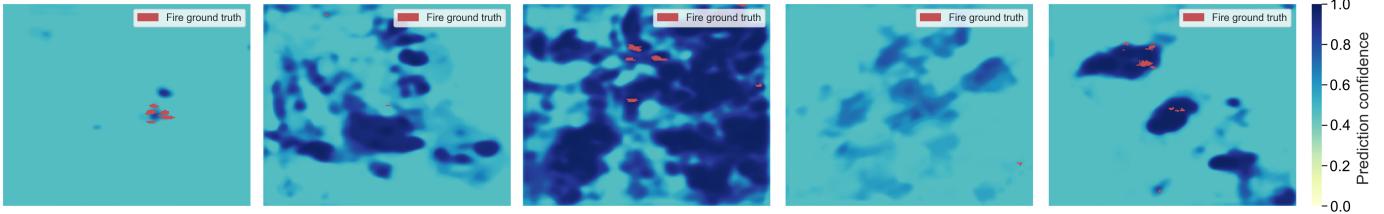


Figure 11: These five pictures are an example of five predictions from the CNN. In red is shown the fire ground truth, while the gradient from white (0.0) to black (1.0) represents the output of the model.

B. CNN variability

10 different CNNs were generated and tested on the validation dataset. The summary of the findings are shown in Figure 12. In order to obtain a representative graph, 10,000 values for each class have been randomly taken out of the results.

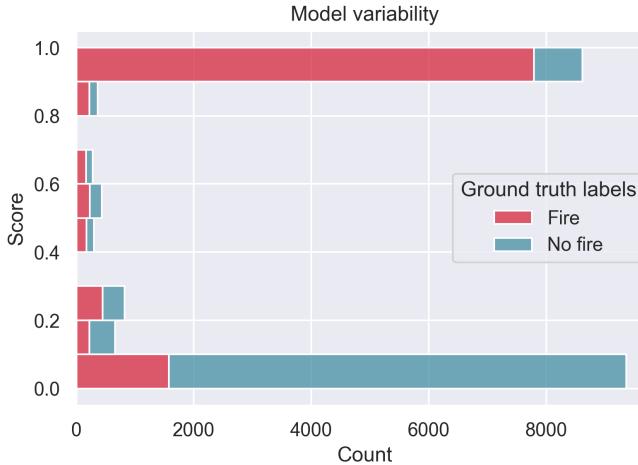


Figure 12: The graph shows the variability of the model on its predictions. In red, the fire instances predicted by the model and in blue the non fire instances predicted.

Looking at the Figure 12, is visible that over all the fire predicted by the model, around 20 % of these are wrongly labelled. The same case is for the no fire instances. In the bar plot, samples are mostly spread between the top and bottom with few instances in the middle.

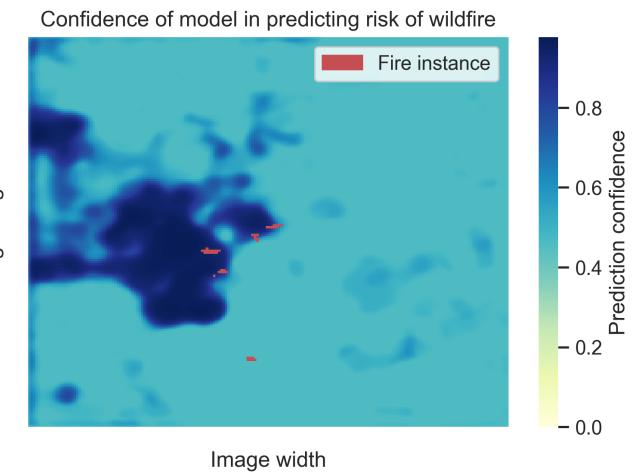


Figure 13: The image is an sample computed by the 10 models. It is calculated using the method illustrated in Figure 10.

Figure 13 shows the calculated spatial variability obtained by 10 models on a single sample from the validation set. The prediction of the models coincide with most of the true ground fire instances. The fire ratio score for this sample is 0.78.

C. Comparing architectures

After optimising the autoencoder and the U-Net, both models showed poor results with the Adam optimiser. Therefore, the SDG optimiser is used for both models. The autoencoder and the U-Net performed best with a learning rate of 0.1 and 0.01, respectively. After their optimisation, all three models are compared in this section and results are shown in Table VI.

Models	Fire ratio	No fire ratio	Dice coefficient	IOU
CNN	0.82	0.87	0.0059	0.0029
U-Net	0.51	0.52	0.0037	0.0018
Autoencoder	0.55	0.60	0.0036	0.0018

Table VI: This table shows results for a CNN, a U-Net and an autoencoder that are trained and evaluated to predict the risk of wildfires.

Looking at table VI, the CNN outperforms the U-Net and the autoencoder achiving a fire ratio of 0.82. The autoencoder has slightly better performances than the U-Net for the fire

ratio and the no fire ratio. This will be further discussed in the next section.

D. Feature analysis

A range of values were added or subtracted to each feature in the validation dataset. This resulted in different outputs from the model. Figure 14 shows results for the fire ratio metric.

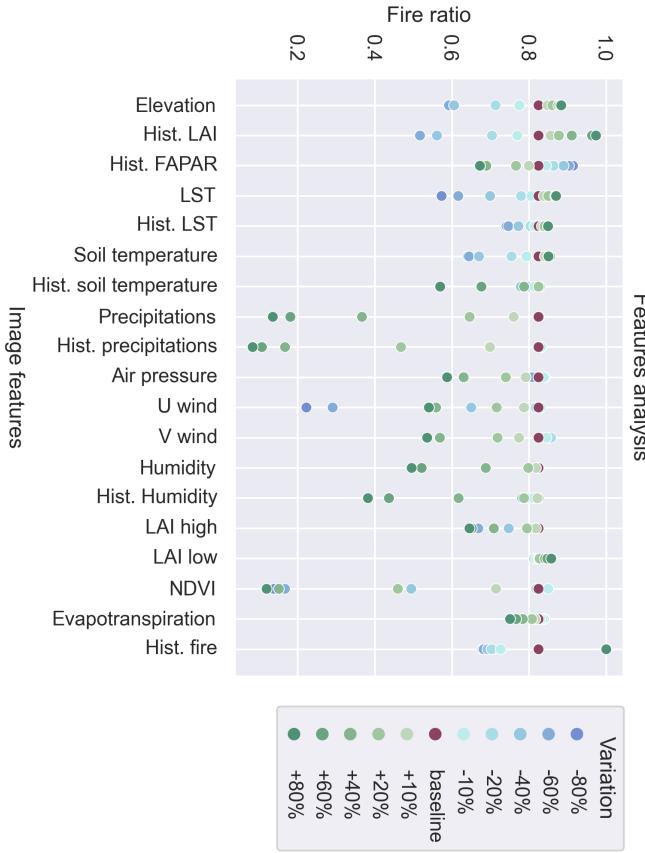


Figure 14: The scatterplot presents the fire ratio results from the variation in each feature. The baseline is the CNN's performance without feature alterations. Positive increments have a green color scheme, while negatives have a blue color scheme.

Looking at Figure 14, a $x\%$ amount increase or decrease for a feature is not equivalent to a $x\%$ amount increase or decrease in the fire ratio. Out of all the features, history of soil temperature, history of humidity, U and V wind component and LAI low and high show strange patterns. These are iteratively removed from the dataset in order to verify their relevance (Figure 17).

The second set of results is shown in Figure 15, while Figure 17 shows the elimination process step by step. Removing u and v components of the wind results in a slightly lower fire ratio score than the baseline. Removing the history humidity feature increases the confidence interval of the results while decreasing its score. Removing the history of soil temperature decreases the fire ratio score of the model. Finally, removing

wind and LAI features increases the fire ratio score. Therefore, these features are not relevant for the fire risk assessment task.

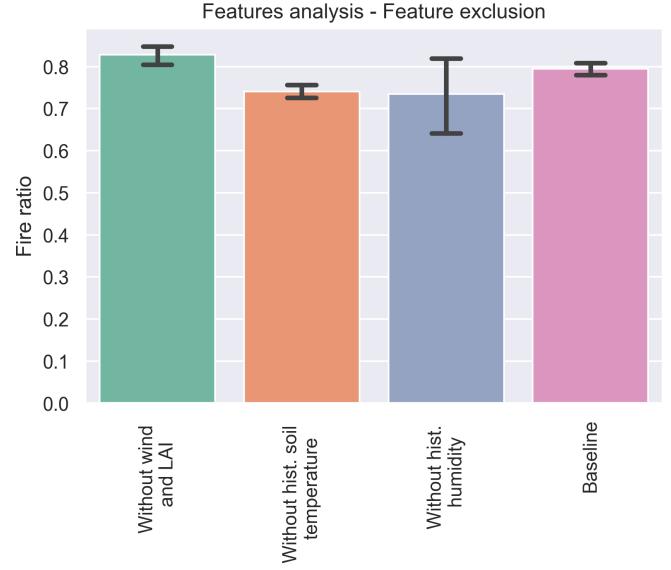


Figure 15: The figure show results for the fire ratio metric by removing some features from the dataset. Each set of tests is run for 5 times and its confidence interval is shown by the line over the bars.

E. Spatial dependence

Six test sets are generated for this experiment. The results for inferring the geographical area of Africa, Asia, Australia, Europe, South America and the USA are shown in Figure 16.

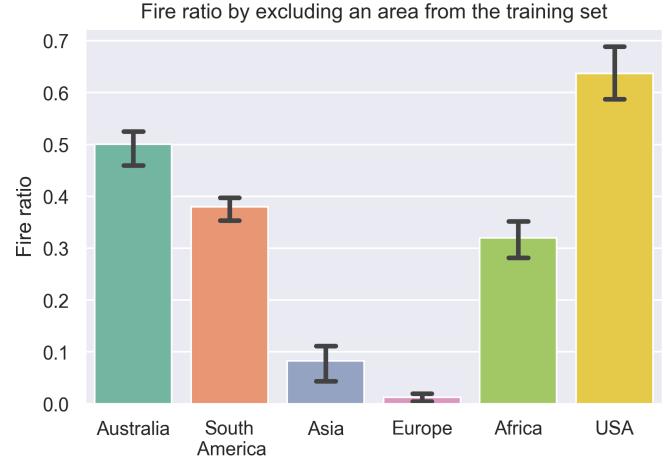


Figure 16: The bar plot shows the fire ratio of different trained CNNs. Each set of tests is run for 5 times and its confidence interval is shown by the line over the bars.

Looking at Figure 16, the score for excluding an area is not uniform across all areas. The best score which is close to previous baseline results, is achieved by inferring on the US area with a score of 0.64. The worst scores are on the Asian

and European areas, 0.8 and 0.1 respectively. Australian, South American and African areas obtain a fire ratio of 0.50, 0.38 and 0.32 respectively.

VI. DISCUSSIONS

A. Data

The samples gathered from GEE do not cover the entire globe either because there is no satellite data available or there have not been fire instances in the area. But mostly, it would convolute the research. One issue with the retrieving images from GEE is the presence of identical fire instances in multiple images. Randomising the sampling could cause wrong results. A better sample selection could be achieved by designing a better GEE algorithm. Images have been manually selected with reference to the date and location of extraction, creating an imbalanced number of samples per areas.

Looking at Figure 11, the model seems to gain a knowledge of where a fire could happen. Fires could be mislabelled due to missing data, fires not started by natural events, etc.

B. CNN optimisation

The exponential learning rate seems to aid the models, achieving consistent results after 30 epochs. Fluctuation is caused by the learning rate being too large, this leads the optimiser into missing the local minima, circulating it. Observing Table V, the CNN with 6 inner blocks and 10 incremental filters per convolutional layer achieved optimal results in predicting the risk of wildfires. This model was more capable of generalising. The higher number of filters results in an increase of the model's complexity, performing slightly better on the training set but not on the evaluation one. The model with 10 incremental filters is used for consecutive tests.

C. CNN variability

Looking at the Figure 12, the CNN's results confirm previous shown results in Table V. The CNNs do not seem to have a high variance over their results. The count of scores that are in between 0.2 and 0.8 relative to the total number of instances is not high (around 2,500). This underlines that the model can label wildfires confidently with some margin of error. These mislabelled fires could also be humanly induced such as accidental or intentional (as an attempt to reduce critical fuel in a forest). The events cannot be predicted or even in most cases, recorded. Furthermore, the low uncertainty presented in the graph emphasises the model's ability to generalise. This factor is essential when building a ML algorithm at global scale.

D. Comparing architectures

Table VI compares the three models over the fire risk assessment task. The CNN performs best by achieving a score of 0.82 and 0.87 for the fire ratio and the no fire ratio, respectively. Comparing the CNN with the autoencoder: the CNN has 16 times more parameters than the autoencoder. This allows the former to learn a bigger number of patterns. Additionally, dropout and batch normalisation layers aid the CNN

to a better regularisation. The U-Net interestingly achieves mediocre results in the fire risk assessment task. This could be due to a lack of dropout layers, although the model already contains batch normalisation layers to improve its regularisation capabilities. The U-Net and the autoencoder use two different layers in their architectures in order to expand back the images. The U-Net implements a *Conv2DTranspose* layer, this is a convolutional layer merged with an upscaling method. Both components have weights to learn. The autoencoder instead makes use of a *UpSampling2D* layer. The latter simply uses a nearest neighbour method to expand back the images. The excessive amount of parameters to learn compared to an autoencoder could be the source of poor performances in the U-Net.

E. Feature analysis

Looking at Figure 14, each feature alteration is discussed here:

- When the *Elevation* increases, the risk of wildfire increments as well. This trend starts to reach a saturation when the vegetation becomes scarce at high altitudes. Additionally, the risk decreases for lower altitudes.
- Negative variation to *Hist. LAI* reduces the fire risk because if there is an absence of fuel there is a lower probability of a fire. Positive variation equals more leafage, which increases the risk of fire.
- Positive increments to *Hist. FAPAR* grow the probability of fire due to plants being dry, while negative variations decrease the fire risk because trees are moistured.
- The risk of wildfires increase with higher *LST*. Although it may be the case for the areas in the dataset, this is not true everywhere in the world (e.g. in deserts and the north and south Pole).
- *Hist. LST* behaves like *LST* but with a lower fluctuation to variations.
- *Soil temperature* acts similarly to *LST*, due to equal characteristics.
- Negative and positive variations do not seem to have a clear trend for the *Hist. soil temperature* feature. Positive increments reduce the risk of wildfire, possibly due to high-temperature areas (e.g. a fire cannot happen in the desert). Negative increments do not give clear feedback. This feature is excluded for redundancy and analysed in the next test.
- With an increase in *Precipitations* there is a lower risk of fire, due to a higher probability of a wet vegetation, reducing the risk of ignition. The decrease in precipitations does not change the output. This could be because the variation leads the feature to reach its minimum level of precipitations or the baseline is already at its minimum. Therefore, the probability of fire is the same.
- Alterations to the *Hist. precipitations* reflect an identical behaviour as the previous feature.
- Positive variations to *Air pressure* result in a lower risk of wildfire, while negative variations reveal an interesting pattern. This behaviour could underline a bias in the

dataset, which is mostly composed of instances between low and medium altitudes, with a few high altitudes. In this case, decreasing air pressure does not change the fire risk due to it already being at its minimum. Increasing the air pressure lowers the fire risk due to the few fire samples at high altitudes.

- *U and V wind* features do not seem to respond well to either positive or negative variations. This feature is excluded for redundancy and analysed in the next test.
- Increasing the *Humidity* lowers the risk of wildfires. Given more moisture in the air, fires are less probable. Negative variations do not create clear feedback on the results. It could be due to a lack of data or that the baseline humidity is at its lowest.
- The *Hist. humidity* seems to reflect the previous feature but the negative variation presents a surprising pattern. This feature is excluded for redundancy and analysed in the next test.
- *LAI high and low* do not respond well to alterations and do not present reasonable patterns. These two are excluded for redundancy and analysed in the next test.
- The *NDVI* feature presents an unexpected response to values more than 10%. This could be because the feature reaches its maximum or minimum, and the model is clueless. For the increase of 10%, the risk of fire is decreased due to water being in the vegetation. While for a decrease of 10%, the risk of fire increases, since the vegetation lacks water.
- *Evapotranspiration* seems to respond well to variation. Positive increments result in a lower risk of fires. This is due to the vegetation being in good condition. Negative increments increase the risk of fires, given drier vegetation.
- *History fire* reveals an interesting pattern. The more fires happened in the area over the past years, the higher the risk of wildfires. Although presumably, after a wildfire there is no much fuel to burn. Looking at the gathered dataset, most fires are extracted around the same hot spot areas, years after years. It may suggest a bias in the data, a lack of fire management or a combination of natural characteristics that make these areas susceptible to wildfires.

Removing history soil temperature and history humidity decreases the ability of the model to predict the risk of fires. Although these features present interesting patterns in the previous test, they seem to be relevant for the problem. Their strange behaviour could be given by irregularities in the data or problems of compatibility with the experiment. The wind components seems to be redundant for this task. The direction of the wind could be for instance more suited to predict where the fire could expand. LAI components are unnecessary due to NDVI, Evapotranspiration, History LAI and History FAPAR features that already give higher spatial resolution insights on the vegetation status. Excluding wind components and LAI features improves the model, gaining a better fire ratio score

over the baseline.

F. Spatial dependence

Studies mentioned in Section I based their work on a single geographical area or generalise their method to the entire globe. This approach is due to multiple dependent and independent variables that contribute to the start of a fire. Looking at Figure 16, it is clear that although there is knowledge inferred, there is also a difference in the characteristics that could cause a wildfire in every area. The CNN can easily predict wildfires in the USA. The latter presents geographical elements in common with some of the other geographical areas. This allows the model to infer with little difficulty. Australia, South America, Africa and the USA have common characteristic patterns. Europe and Asia wildfires do not seem to be inferred as easily as other areas. This could be because wildfires are managed differently (e.g. leaving no time for data logging) or the characteristics of a fire differ from other areas. The CNN achieves a fire ratio of 80% on the validation set. The latter is mostly composed of samples from the US area. Looking at the spatial dependence results, these areas seem to be easily predicted by the model. This dataset split was not intentional. An alternative area combination for the validation set should cause a lower fire ratio score. Lastly, building different models for different areas around the world could be as optimal as a *global* model. One instance in which this statement is not true is the following example. Given climate change and rising temperature, new areas around the globe are starting to present wildfires. These areas do not have enough wildfire data. A *global* model could be easily deployed over these areas achieving optimal results.

VII. CONCLUSION

The evolution of machine learning techniques and the easy availability of satellite data create a perfect combination to better develop models able to predict extreme events and potentially save catastrophes.

In this research, a CNN capable of assessing a wildfire risk over the entire globe is built. This is achieved by building a dataset with multiple features and feeding this to the algorithm. Furthermore, this architecture is light enough to be integrated in a near real-time pipeline while achieving a resolution of 1000m/pixel.

Gathering data from different areas around the world and with multiple features is challenging. Special effort has been invested in creating training and validation set with independent samples.

The CNN architecture achieves a good trade-off between the score of the four metrics and the number of weights to learn. Various parameters have been used for the optimisation task. As experiments show, this type of architecture is better suited than a U-Net or an autoencoder for risk assessment of wildfires. The U-Net architecture tested here showed poor results. A different one could give better results.

Experiments on the variability of the model shows it can correctly classify a risk of fire around 80% of the time.

Uncertainty in the model could be reduced by increasing the number of training samples. Mislabelled fire instances could be related to external and independent events.

Features are easily analysed by studying their response to positive or negative variations on the validation set. Out of 20 features, 4 are removed due to redundancy. Deleting features improves the model. These experiments have not been done by training the model excluding features iteratively. Therefore, different methods could have different results.

Each area around the world has different geographical characteristics. This makes building a robust global algorithm very challenging. Inferring knowledge in some areas seems to be easier than in others, mostly because of similar characteristics or because there is more clear data available.

Overall, this research has shown that a simple CNN can be successful in predicting the risk of wildfires in areas around the globe. Improving this technique and using it on a larger scale could save damage to the economy, properties, wildlife and humans. This especially holds true in areas with poor wildfire management. Although there are some drawbacks regarding the availability and frequency of the data, satellite data is an essential resource for this and other similar applications. An attempt to deploy the model on GEE has been made, but given time constraints and the high cost, it has been dismissed.

Future work will focus on:

- improving the GEE algorithm for extracting samples such that there is no overlap between samples;
- exploring higher temporal and spatial resolution sources, for instance *PlanetScope Dove*. It provides daily high resolution images (between 3m and 5m/pixel). High costs are expected with this type of workflow;
- comparing multiple loss functions (e.g. Tversky Loss, Lovasz Hinge Loss, Combo Loss);
- further investigating the feature space of the dataset by first iteratively deleting a feature from the dataset and then training the model;
- inspecting in more details the reason why the model mislabels a fire instance as a no fire instance;
- deploying the model on GEE.

REFERENCES

- Andela, Niels et al. (2019). “The Global Fire Atlas of individual fire size, duration, speed and direction”. In: *Earth System Science Data* 11.2, pp. 529–552.
- Artés, Tomàs et al. (2019). “A global wildfire dataset for the analysis of fire regimes and fire behaviour”. In: *Scientific data* 6.1, pp. 1–11.
- Chen, Gongbo et al. (2021). “Mortality risk attributable to wildfire-related PM_{2.5} pollution: a global time series study in 749 locations”. In: *The Lancet Planetary Health* 5.9, e579–e587.
- Coen, Janice L and Craig C Douglas (2010). “Computational modeling of large wildfires: a roadmap”. In: *2010 Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science*. IEEE, pp. 113–117.
- Crippen, R et al. (2016). “NASADEM global elevation model: methods and progress”. In: *Evapotranspiration and the Water Cycle* (2018). <https://www.usgs.gov/special-topics/water-science-school/science/evapotranspiration-and-water-cycle>.
- Foundation, First Street (2022). *Fueling the Flames, The 5th National Risk Assessment: Fueling the Flames*. <https://firststreet.org/research-lab/published-research/article-highlights-from-fueling-the-flames>.
- Huot, Fantine et al. (2021). *Next Day Wildfire Spread: A Machine Learning Data Set to Predict Wildfire Spreading from Remote-Sensing Data*. arXiv: 2112.02447 [cs.CV].
- Jadon, Shruti (2020). “A survey of loss functions for semantic segmentation”. In: *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, pp. 1–7.
- Jazebi, Saeed, Francisco De Leon, and Albert Nelson (2019). “Review of wildfire management techniques—Part I: Causes, prevention, detection, suppression, and data analytics”. In: *IEEE Transactions on Power Delivery* 35.1, pp. 430–439.
- Justice, CO et al. (2002). “The MODIS fire products”. In: *Remote sensing of Environment* 83.1-2, pp. 244–262.
- Martín Abadi et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. URL: <https://www.tensorflow.org/>.
- Mizutori, M and Debarati Guha-Sapir (2017). “Economic losses, poverty and disasters 1998–2017”. In: *United Nations Office for Disaster Risk Reduction* 4, pp. 9–15.
- National Interagency Fire Center (NIFC) (2021). <https://www.nifc.gov/fire-information/statistics>.
- Nieradzik, Lars et al. (2021). “Effect of the output activation function on the probabilities and errors in medical image segmentation”. In: *arXiv preprint arXiv:2109.00903*.
- Park Williams, A et al. (2013). “Temperature as a potent driver of regional forest drought stress and tree mortality”. In: *Nature climate change* 3.3, pp. 292–297.
- Radke, David, Anna Hessler, and Dan Ellsworth (2019). “FireCast: Leveraging Deep Learning to Predict Wildfire Spread.” In: *IJCAI*, pp. 4575–4581.
- Rahman, Md Atiqur and Yang Wang (2016). “Optimizing intersection-over-union in deep neural networks for image segmentation”. In: *International symposium on visual computing*. Springer, pp. 234–244.
- Rothermel, Richard C (1972). *A mathematical model for predicting fire spread in wildland fuels*. Vol. 115. Intermountain Forest & Range Experiment Station, Forest Service, US ...
- Sayad, Younes Oulad, Hajar Mousannif, and Hassan Al Moatassime (2019). “Predictive modeling of wildfires: A new dataset and machine learning approach”. In: *Fire safety journal* 104, pp. 130–146.
- Stein, Susan M et al. (2013). “Wildfire, wildlands, and people: understanding and preparing for wildfire in the wildland-urban interface—a Forests on the Edge report”. In: *Gen. Tech. Rep. RMRS-GTR-299. Fort Collins, CO. US Department*

of Agriculture, Forest Service, Rocky Mountain Research Station. 36 p. 299.

Surya, Lakshmisri (2020). "Fighting fire with ai: Using deep learning to help predict wildfires in the us". In: *International Journal of Creative Research Thoughts (IJCRT)*, ISSN, pp. 2320–2882.

US EPA (2021). <https://www.epa.gov/climate-indicators>.

Wells, Gail (2008). "The Rothermel Fire-Spread Model: still running like a champ". In.

What is NDVI (2022). <https://gisgeography.com/ndvi-normalized-difference-vegetation-index/>.

What is remote sensing? U.S. Geological Survey (USGS) (2021). <https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used>.



Figure 18: The image shows the location of the samples that compose the training and validation sets for the spatial dependence tests.

VIII. APPENDIX

	Instances type	Instances
Training	no fire	35,927,786
	fire	72,214
Validation	no fire	5,988,466
	fire	11,534

Table VII: Dataset classes distribution.

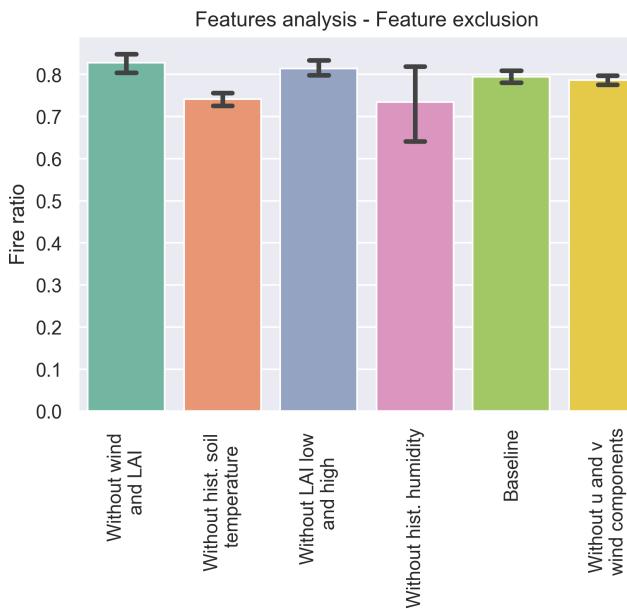


Figure 17: The figure show the complete results for the fire ratio metric by removing some features from the dataset. Each set of tests is run for 5 times and its confidence interval is shown by the line over the bars.