Time-Driven Fire Risk Forecasting: Leveraging Historical Trends for Enhanced Seasonal Modeling

Preprint · July 2023

DOI: 10.13140/RG.2.2.28395.31528/1

Roma Jain
Ambee

Akshath Raghav Ravikiran
Purdue University

Pareekshith Katti
Ambee

Time-Driven Fire Risk Forecasting: Leveraging Historical Trends for Enhanced Seasonal Modeling

Roma Jain, Akshath R Ravikiran, Pareekshith US Katti

Abstract

Wildfires pose severe threats across North America, causing extensive damage to lives, ecosystems, and property. To address this, accurate fire prediction and forecast outlooks are crucial for effective mitigation. Agencies like the National Interagency Fire Center (NIFC) and the Canadian Wildland Fire Information System (CWFIS) provide vital fire risk assessments. In this paper, our main goal was to demonstrate the sufficiency of historical fire risk data for accurate forecasting. We focused on weather-calculation-based fire risk prediction models, specifically exploring the temporal aspect's importance in enhancing accuracy. Two encoding methods, One-Shot and Year-By-Year, used for encoding the seasonal changes of fire weather, were analyzed for their implications in fire risk assessment, revealing contrasting attributes. The One-Shot model shows superior accuracy and favorable plots, while the Year-By-Year model offers alternative insights. Despite minor differences in feature importance between the two models, both effectively utilized historical fire weather data for forecasting. This study contributes significantly to fire risk prediction by providing a comprehensive analysis of temporal influences and the effectiveness of different encoding methods. The findings guide model design improvements, bolstering wildfire management and protection measures.

Introduction

The prediction and estimation of fire risk season have become increasingly crucial due to the alarming impact of wildfires on human lives and the environment. Recent fire incidents have resulted in tragic loss of life, such as the Black Summer of 2019-2020 in Australia, where an estimated 33 people perished, and extensive damage to land and homes, with 18.6 million hectares consumed by flames (Parrott et al., 2021). In early 2021, wildfires ravaged approximately 7.13 million acres in the United States (Statista, 2022), while Algeria experienced a prolonged system of wildfires that took the lives of at least 70 people (Reuters, 2021). The financial toll is equally significant, with the 2018 California Camp Fire alone incurring losses exceeding $16.5 billion (Los Angeles Times, 2019). These losses are primarily attributed to the destruction of residential and commercial structures, including homes, businesses, and infrastructure. The costs also encompassed expenses related to emergency response efforts, rehabilitation, and restoration of the affected areas. In the coming days, resources from all over the Western United States would be sent to Paradise to assist in the fight to contain the fire. In the aftermath of the wildfire, resources from across the Western United States were mobilized to support firefighting efforts in Paradise (Town of Paradise, n.d.). By November 10th, the response included a force of 5,596 firefighters, 622 engines, 75 water tenders, 101 fire crews, 103 bulldozers, and 24 helicopters working diligently to contain the fire (Town of Paradise, n.d.). These statistics highlight the urgent need for comprehensive fire risk forecasting and management strategies to protect lives, infrastructure, and ecosystems.

Understanding the seasonality of fire is crucial for proactive fire management. Climate change, characterized by rising temperatures, changes in precipitation patterns, and prolonged droughts, has been identified as a significant driver of increased fire risk.

Additionally, human activities, such as urbanization and land-use changes, have contributed to the expansion of the wildland-urban interface, where residential areas intersect with fire-prone landscapes, thereby increasing the vulnerability of communities to wildfires. The combination of these factors has resulted in a rise in the frequency, intensity, and extent of wildfires in many regions globally.

While extensive research has been conducted on the weather conditions that contribute to wildfire occurrence (Sihan Li et.al., 2021), the focus on historical fire season trends and their specific implications in fire forecasting has been relatively limited. Previous studies have primarily explored the relationship between fire and individual weather variables or examined long-term climate trends. However, a more nuanced understanding of how fire patterns vary within different seasons and how they interact with other temporal and spatial factors remains to be addressed.

The main contribution of this paper is twofold: firstly, assessing the efficacy of historical data trends for fire season forecasting, and secondly, analyzing the results of fire season forecasts using forecasted weather conditions. By investigating the role of historical data trends and the potential benefits of integrating forecasted weather data, this research aims to advance the field of fire season prediction and provide insights that can inform proactive fire management strategies.

Delving deeper into the seasonality of fire enables researchers to uncover valuable insights into the specific temporal dynamics of fire occurrence, including the timing, duration, and intensity of fire seasons. Such knowledge is essential for developing accurate forecasting models. Understanding the region-specific cyclicity of fire is crucial for regional forecasting accuracy. Fire occurrences exhibit distinct patterns that vary across different regions, influenced by unique combinations of weather conditions and vegetation growth. Narrowing our focus to the intricacies within unique geographic boundaries, like ecoregions, biomes, and continents, will help us gain valuable insights into the variations in fire behavior that are specific to each area and by incorporating different target encoding methods into machine learning models formulated for our study, we aim to capture the seasonal patterns in fire risk.

Literature Review

The literature review of our paper delves into various studies focusing on fire weather seasons and wildfire risk prediction across different regions. One recurring observation is the cyclical nature of fire seasons in most regions, with a notable upward trend in the length of fire seasons in recent years. For instance, in Australia, the research by S. Harris Lucas and Chris Lucas (2019) highlights the strongest wildfire risk seasonality in the spring months in the South East. Similarly, in South America, Yang Chen and James Randerson (2011) explore the relationship between sea surface temperature anomalies and fire weather season. To forecast fire weather seasons, studies like Alan Mandal and Adam Kochanski (2022) and Xiao-Rui Tian and Xue-Zheng Zong (2020) utilize forecasted weather conditions and ENSO, respectively. Several other related papers also employ methods that build on forecasted weather conditions.

On the other hand, certain studies, like the one by J. Bedia, S. Herrera, D. San Martín, N. Koutsias & J. M. Gutiérrez (2013), rely on historical fire weather index drivers such as wind speed, humidity, temperature, and precipitation. Calheiros T, Pereira MG, Nunes JP, et al. (2021) assess the evolution of fire weather indices and the Number of Extreme Days (NED) in the context of climate change, suggesting potential changes in fire weather patterns. Hantson et al. (2016) conduct a comparative study on how fires are represented in fire-enabled dynamic global vegetation models. Miller et al. (2023) utilize the Canadian regional climate model version 5 (CRCM5-LE) to reveal increasing fire danger trends in Central Europe.

Additionally, Kátia Fernandes, Michael Bell, and Angel G Muñoz (2022) combine the Subseasonal Experiment (SubX) ensemble mean precipitation forecast and a vegetation health index (VHI) to assess fire risk in the Amazon, resulting in enhanced prediction accuracy. Richardson et al. (2022) analyze the relationship between drought and fire weather, finding that regions prone to wildfire disasters are more likely to experience fire weather years preconditioned by drought. San-Miguel-Ayan J., Houston-Durrant T, et al. (2022) provide an overview of the fire season of 2022 in Europe, the Middle East, and North Africa, assessing the impact of wildfires and fire danger conditions.

By examining these diverse studies, we aimed to gain a comprehensive understanding of fire weather season dynamics and enhance the accuracy of wildfire risk prediction in various regions. The research conducted in these studies plays a crucial role in advancing our knowledge of fire behavior and providing valuable insights for developing effective wildfire management strategies tailored to different geographical locations.

Methodology

In this section, we describe the methods and data used in our paper, which aims to assess the efficacy of historical data trends for fire season forecasting. We explain the formulation of the problem, the target variable definition, present the list of features used, and visualize the same for our understanding. Additionally, we discuss the approaches employed for target encoding and the models used for fire season risk prediction.

The motivation behind our study design stems from a compelling need to thoroughly analyze the performance of our models, which are solely trained on historical spatiotemporal data. To achieve an extensive evaluation, we will compare the forecasts generated by our models with those produced by multiple-member ensembles. These ensembles utilize monthly weather prediction models and leverage climate-derived weather formulas to enhance forecast accuracy. By contrasting our model's predictions with the outcomes generated by these advanced weather-based ensembles, we aim to gain valuable insights into the model's effectiveness, robustness, and its ability to make accurate fire risk forecasts under various climatic conditions.

The problem formulation revolves around predicting fire season risk using the historical FWI values from the gridded North American continent. The dataset utilized in our research is sourced from the Global Fire Weather Database (GFWED), specifically obtained from the GEOS-5 model provided by the Global Modeling and Assimilation Office. This comprehensive data source offers historical Fire Weather Index (FWI) values at a spatial resolution of 0.25x0.25 (in degrees), enabling a detailed assessment of fire risk patterns across various regions. By leveraging this dataset, which combines advanced weather modeling and assimilation techniques, we conducted a thorough analysis of fire risk dynamics and enhanced the accuracy and reliability of our model's predictions.

The historical FWI data collected spans a period of eight years, with a daily temporal resolution. The aggregation methodology calculates the average FWI value for the month and encodes it as a 'fire_risk' value. The target variable, 'fire_risk', is defined based on the severity of fire risk, categorically labeled into 'No Risk', 'Low Risk', 'Moderate Risk', 'High Risk', and 'Extreme Risk'. Then we derive monthly percentage values for each of these risk levels, which depict the percentage of the month for each geo-location that is assigned one of the above levels.

In our research, we integrate the gridded GEOS data with ecoregions and biomes overlays to assign specific land type identifiers to all geolocations. Ecoregions are distinct ecological regions characterized by their unique climate, geology, and vegetation, encompassing similar ecological characteristics across geographic areas. Biomes, on the other hand, represent large-scale ecological communities dominated by specific plant and animal species, often influenced by factors such as temperature, precipitation, and soil conditions.

By incorporating Google Earth's Ecoregions and Biome data (Google Earth Engine, n.d.), we categorized regions based on their unique characteristics, including land type and fuel type. This crucial step significantly impacts the prediction of fire risk, allowing us to tailor wildfire management strategies to specific ecological settings. The integration of this essential data enhances the accuracy and granularity of our model's fire risk predictions, enabling more effective wildfire prevention and response efforts that are tailored to the specific land and fuel types within each region.

During our data exploration stage, we examined the distribution of historical fire weather indexes in North America. To visualize the extent of high-fire-risk areas across the entire North American region, we plotted the distribution plot for the monthly percentage. The results demonstrated a gradual increase in the fire-prone area over the last 8 years, with the 2022-2023 period showcasing the most pronounced and gradual rise in fire risk throughout the annual fire seasons.

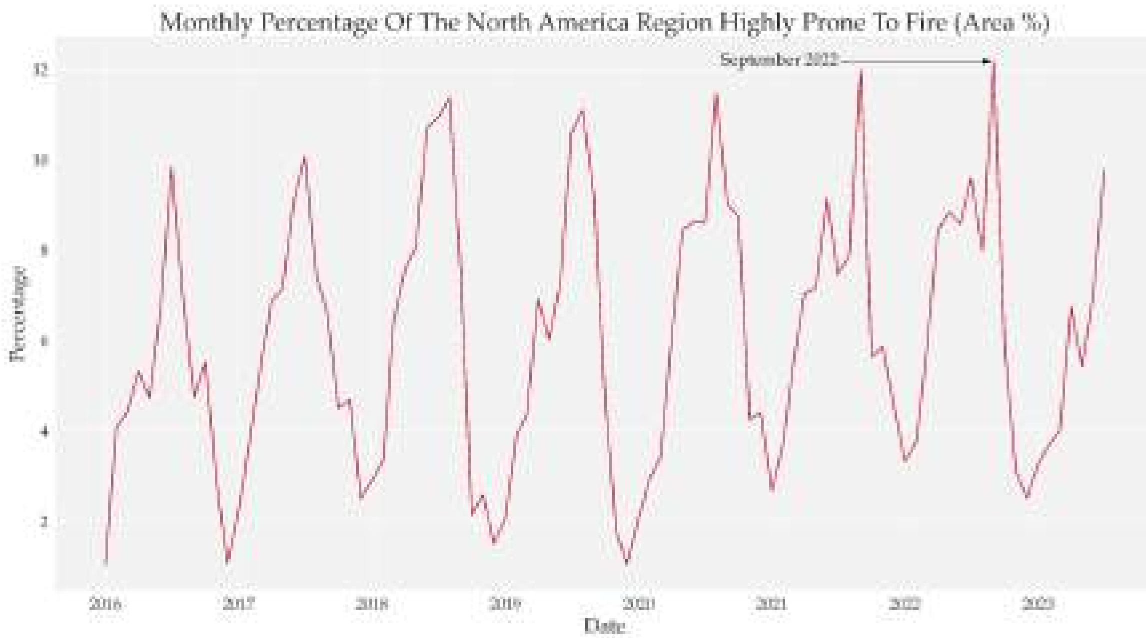

Fig 1: Time series plot illustrating the monthly variation in highly prone fire areas as a percentage of North America's total land area

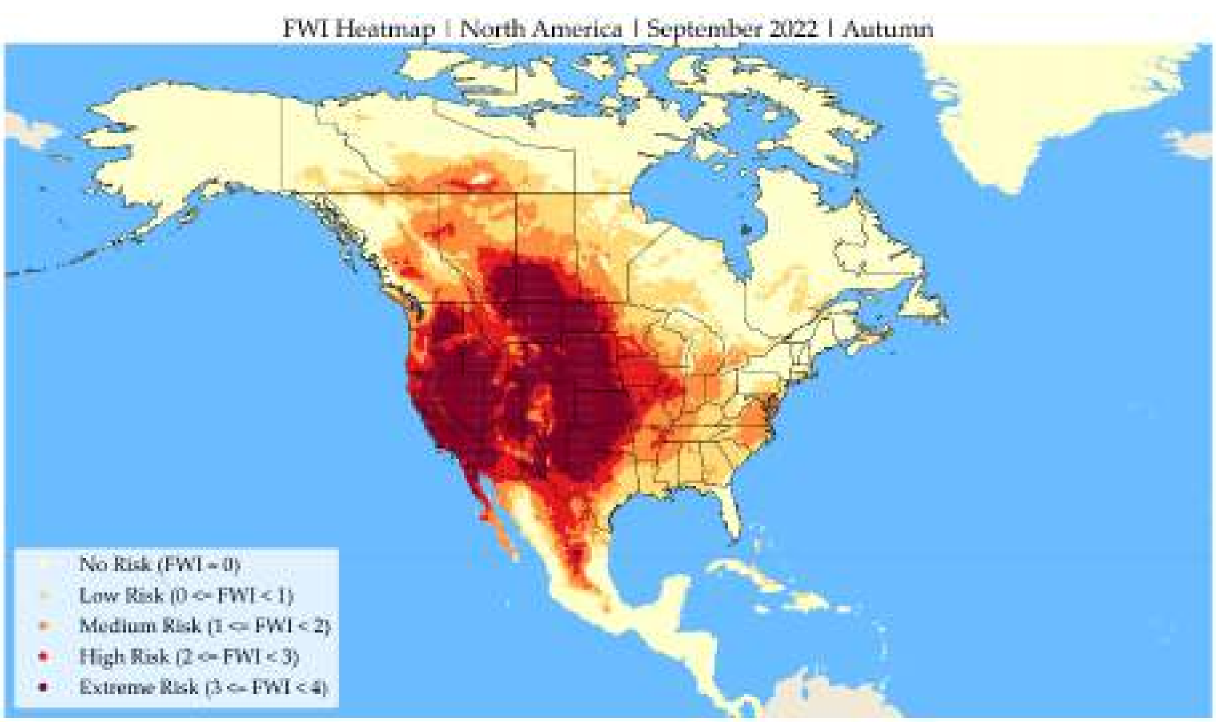

Fig 2: FWI Heatmap of North America for September 2022

Our objective is to develop models capable of accurately predicting fire season risk by leveraging the spatiotemporal dynamics and shifts in fire risk patterns across different years hence the features curated for capturing spatiotemporal aspects are:

1. 'geo': This feature represents the geographical location of the fire risk data point from the gridded GMAO FWI data. It provides information about the grid cell associated with each data point.

2. 'BIOME_NUM': This feature represents the biome number associated with the fire risk data point. Biomes are major ecological communities characterized by distinctive vegetation and climate patterns.

3. 'ECO_ID': This feature represents the ecoregion ID associated with the fire risk data point. Ecoregions are areas with similar ecological characteristics and environmental conditions.

4. 'fire_risk': This feature represents the target variable, indicating the level of fire risk associated with each data point. It ranges from 'No Risk' (0) to 'Extreme Risk' (4), capturing the severity of fire danger.

5. 'no_risk_percentages', 'low_risk_percentages', 'moderate_risk_percentages', 'high_risk_percentages', 'extreme_risk_percentages': These features represent the percentages of data points falling into different fire risk categories (no risk, low risk, moderate risk, high risk, extreme risk). They provide insights into the distribution and prevalence of fire risk levels in the month used for training the model to forecast for the next month.

6. 'no_risk_percentages_(n-11)', 'low_risk_percentages_(n-11)', 'moderate_risk_percentages_(n-11)', 'high_risk_percentages_(n-11)', 'extreme_risk_percentages_(n-11)': This feature is to represent fire risk levels as

the percentage of the days in the previous year's month. They capture the Year-By-Year shifts and trends in fire risk levels.

7. 'year', 'quarter': These features represent the temporal information associated with each data point, such as the year and the quarter. They enable the models to capture the temporal dynamics and variations in fire risk patterns across different time periods.

8. 'month_encoded_no_risk', 'month_encoded_low_risk', 'month_encoded_moderate_risk', 'month_encoded_high_risk', 'month_encoded_extreme_risk': These features encode the months into numerical representations corresponding to different fire risk levels looking at all the years considered in the training data. They capture the seasonal variations in fire risk. We will dive deeper into this in the later half of this section.

9. 'target_month_encoded_no_risk', 'target_month_encoded_low_risk', 'target_month_encoded_moderate_risk', 'target_month_encoded_high_risk', 'target_month_encoded_extreme_risk': These features encode the target month into numerical representations corresponding to different fire risk levels looking at all the years considered in the training data. They capture the seasonal variations in fire risk.

Fig 3: Distribution of FWI values, averaged monthly, across the entire training dataset for North America

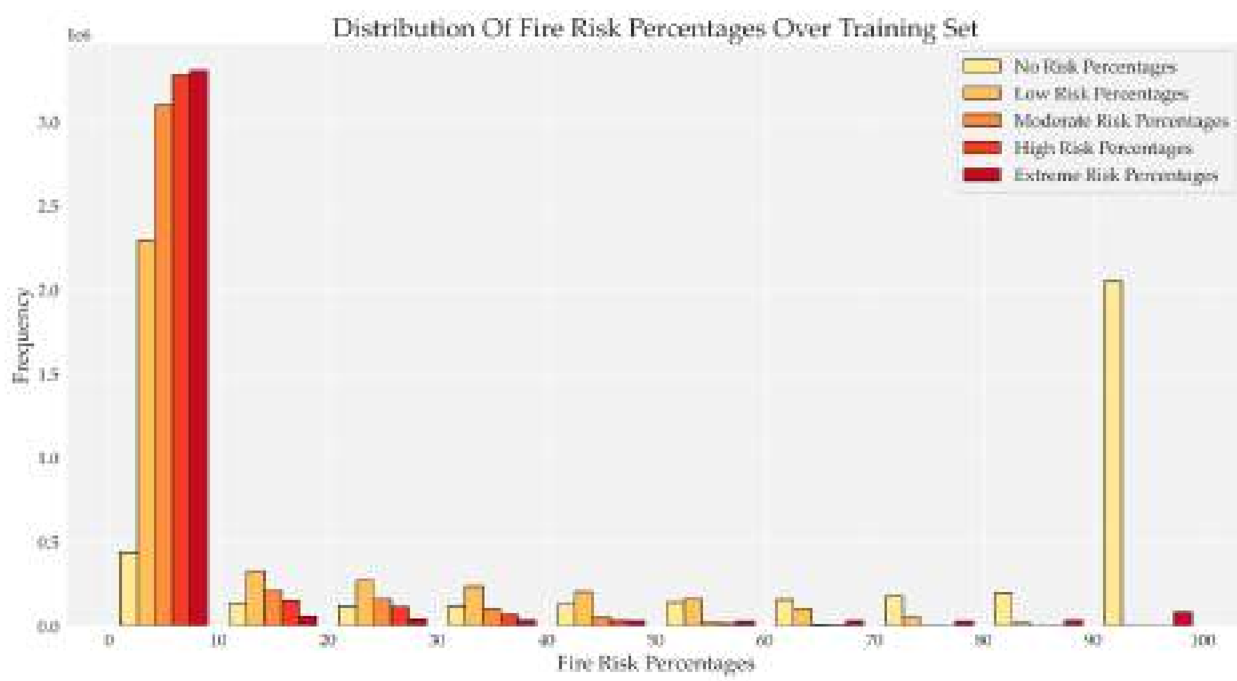

Fig 4: Distribution of FWI percentages, categorically averaged monthly, across the entire training dataset

To account for year-wise patterns and shifts in fire risk, we employ two target encoding approaches: One-Shot encoding and Year-By-Year encoding of the 'month' feature across the 'percentage' features. These encoding techniques enable our models to capture the temporal dynamics and variations in fire risk, enhancing their predictive capabilities. One-shot encoding calculates the average risk level for each month across all previous years, while Year-By-Year encoding computes the expanding average risk level for each month considering all previous years. These encoding methods are implemented within the timeframe of values utilized during model training in time-series-based folds to avoid data leakage. Given the observed prolongation of fire seasons in previous years, it is crucial to ensure that the model does not have access to data from subsequent months while predicting fire risk for the following month to avoid data leakage.

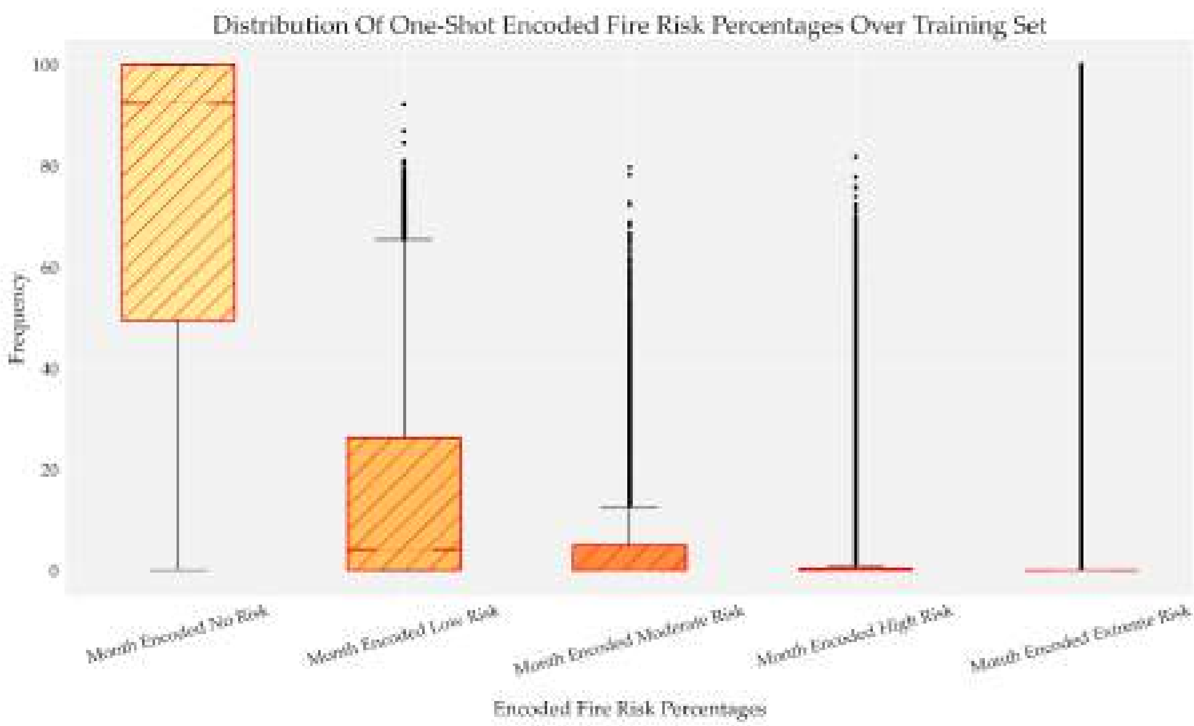

Fig 5: Boxplot showing the distribution of fire risk percentages, encoded using One-Shot encoding, over the training dataset

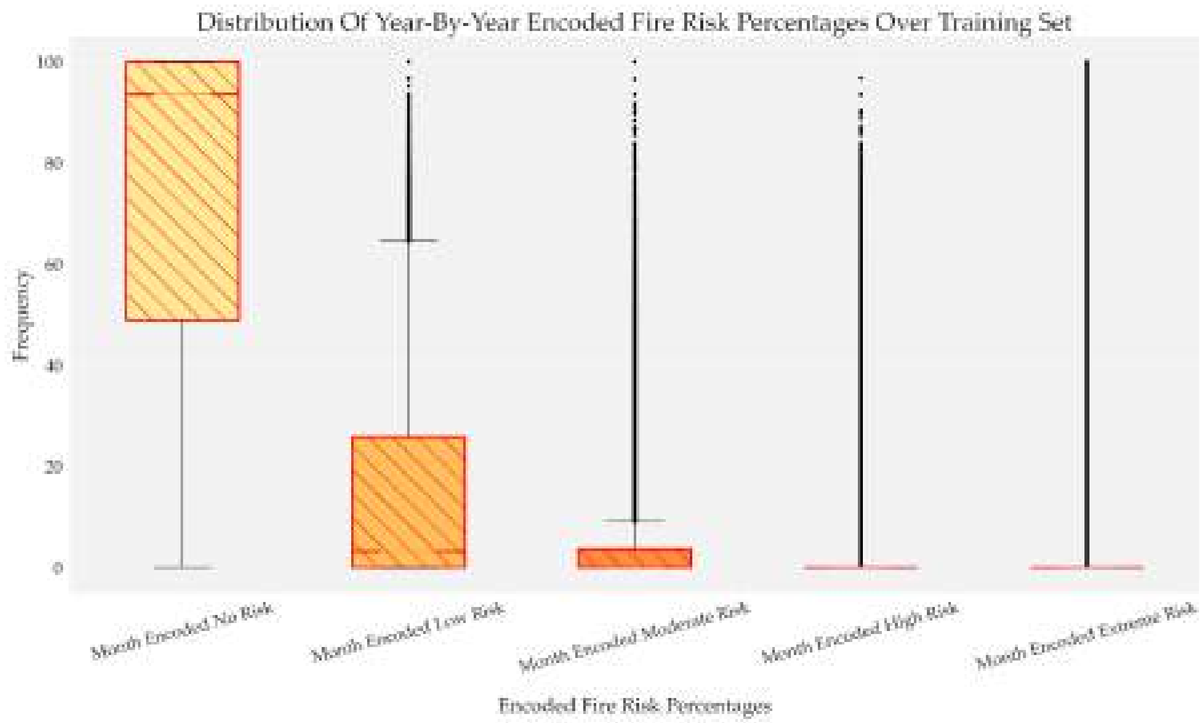

Fig 6: Boxplot showing the distribution of fire risk percentages, encoded using Year-By-Year encoding, over the training dataset

Upon comparing the boxplots of One-Shot and Year-By-Year encoding methods, we observe that the distribution of month-encoded values for No Risk and Extreme Risk is similar, while Low Risk, Moderate Risk, and High-Risk values exhibit lower outliers in the One-Shot plot compared to the Year-By-Year plot. This difference is further highlighted by the smaller interquartile range in the Year-By-Year plot for the same three risk values. The nature of the Year-By-Year encoding method becomes evident when considering a geo-location that incrementally experienced extreme fire risk over the month of July across different years. With Year-By-Year encoding, the risk level values assigned to the month of July vary based on previous years, capturing changes in the fire risk trend. In the early years, these values would be lower but would go on increasing sharply during the later years. In contrast, the One-Shot encoding assigns the same values to the month of July across all years, potentially overlooking the historical changes in fire risk trends. Despite this, our results, discussed later on, showcase the good accuracy achieved by both encoding methods, indicating the overall season's reliability in modeling fire risk historically.

For fire season risk prediction, we employed the Random Forest algorithm in Python. It was selected based on its capability to handle complex relationships and capture nonlinear patterns inherent in the data. To determine the optimal combination of parameters, we utilized cross-validation techniques over the training dataset, adjusting key parameters such as n_estimators (number of decision trees), max_depth (maximum depth of the trees), min_samples_split (minimum number of samples required to split an internal node), min_samples_leaf (minimum number of samples required to be at a leaf node), and max_features (number of features to consider when looking for the best split).

In summary, our methodology encompasses several key steps. We begin by analyzing the historical gridded Fire Weather Index (FWI) values spanning multiple years. We enhance the dataset by incorporating temporal and spatial features that capture the spatiotemporal dynamics of fire risk. To further enrich the predictive capabilities of our models, we apply One-Shot and Year-By-Year target encoding techniques, which enable our chosen RandomForestRegressor models to capture the temporal aspects of fire season risk.

Before we present our results, we will delve into our custom time-series-based 'N-Fold' training process, a vital aspect of our model's development and validation. We iterate through a series of time steps with the following calculations in a similar process with the goal of achieving reproducibility of the same.

The folding process involves multiple steps for each fold. In Fold 1, we take the current month as June 2022, and the forecast month as July 2022. The recalculation set for this fold spans from January 2016 to June 2022 and includes the current month's data to encode values with real-time information during training. The training set encompasses data from January 2017 to May 2022 and is used to train the model with encoded values representative of the current month's data. The testing set contains data from June 2022, providing encoded values representing the model's exposure to data in real time. Additionally, the validation set consists of data from July 2022, serving as an unbiased test to evaluate the model's ability to handle the unpredictable nature of wildfire events, as this data will not be able to influence the model's parameters.

Similarly, in Fold 2, the current month is July 2022, and the forecast month is August 2022. The recalculation set spans from January 2016 to July 2022, including the current month's data for real-time encoding. The training set contains data from January 2017 to June 2022, allowing the model to train with encoded values representative of the current month, now July 2022. The testing set includes data from July 2022, providing encoded values that represent the model's exposure to historical data in real time. Finally, the validation set consists of data from August 2022, which will serve as the ground truth while comparing against the values predicted from the test set. This process continues for subsequent folds.

Our main motive in doing this is to ensure that the model is learning one month into the past from the current month and retain some data to test and record the way the model captures the changing nature of data. Moreover, this method prevents any possible leakage of data from the past into the future, especially during the re-calculation of encoded values.

The proposed methodology aims to forecast the percentage of the coming month which will be under certain risk levels. From these predicted risk values, we have derived a monthly-averaged fire risk forecast, which represents the risk that a particular location would be under for the majority of the month.

The subsequent sections of this research paper will delve into the specific details and present the results obtained from our comprehensive analysis. We will discuss the methods we have used to evaluate the performance of our models, the model performance, the impact of different encoding techniques, and provide insights into the effectiveness of historical data trends in accurately predicting fire season risk.

Results

As we have explained in the earlier sections, our historical FWI values vary in the range of 0 to 4, with each real integer representing a risk level. The same applies to the derived monthly-averaged forecast value that we discussed above. For our understanding, we classify No Risk (0) and Low Risk (1) into 'Negative' class and High Risk (3) and Extreme Risk (4) as 'Positive' Class, to denote the possibility of fire occurrences. Similarly, we choose error metrics that help us understand whether the model, in broad terms, is able to classify geolocations into one or the other class. For this reason, we measure accuracy within an error range of a single risk level, and call it 'Weighted-average Accuracy'. We average out the count of geo-locations where the model has predicted the monthly risk factor within the error range of one risk level. On the other hand, geolocations, where the model has been highly overestimated or highly underestimated, have been solely penalized. This affects only those locations which the model classifies into a 'Positive' risk class where, in fact, there might be a 'Negative' possibility of fire.

We also record the accuracy of every predicted risk percentage target variable, in a similar manner. A value of 10 predicted represents the percentage of a month which is, approximately, 3 days. We interpret a difference of less than 10 (percent) between the real and predicted values to be an acceptable fire risk prediction. Here, we take the weighted average of the geolocations where the model predicts a higher than 10 (percent) error difference with the real values as the 'Percentage Accuracy'.

In the ensuing segments, we shall present a comprehensive exposition of the error metrics pertaining to our developed models. Additionally, we shall provide illustrative plots that juxtapose the actual observed data against the corresponding predicted data for each fold within our training process. Through these analyses, we aim to gain valuable insights into how effective our models are in dealing with the complexities and uncertainties in the subject area.

First, we will provide a brief about the fire seasons in the North America Region and the parts of the continent that are generally placed under fire-prone classifications throughout the year. During the winter months of December, January, and February, fire activity is concentrated in the SouthWest and South areas. Spring, spanning March, April, and May, sees fire season expanding to cover the North, Central, and SouthWest regions. As summer arrives in June, July, and August, wildfire activity becomes prevalent in multiple regions, including the Northwest, South, Southwest, and East areas. Lastly, autumn, encompassing September, October, and November, witnesses the fire season primarily concentrated in the Central region. Additionally, California, due to its vulnerable ecosystem and climate conditions, is now experiencing a year-round fire risk (Cart, 2022, para. 3). We draw these conclusions by analyzing the following region map and information recorded about fire climate regions (NWCG, n.d.). Presented below are visualizations of the fire climate regions in North America, depicting their spatial distribution and corresponding fire risk levels.

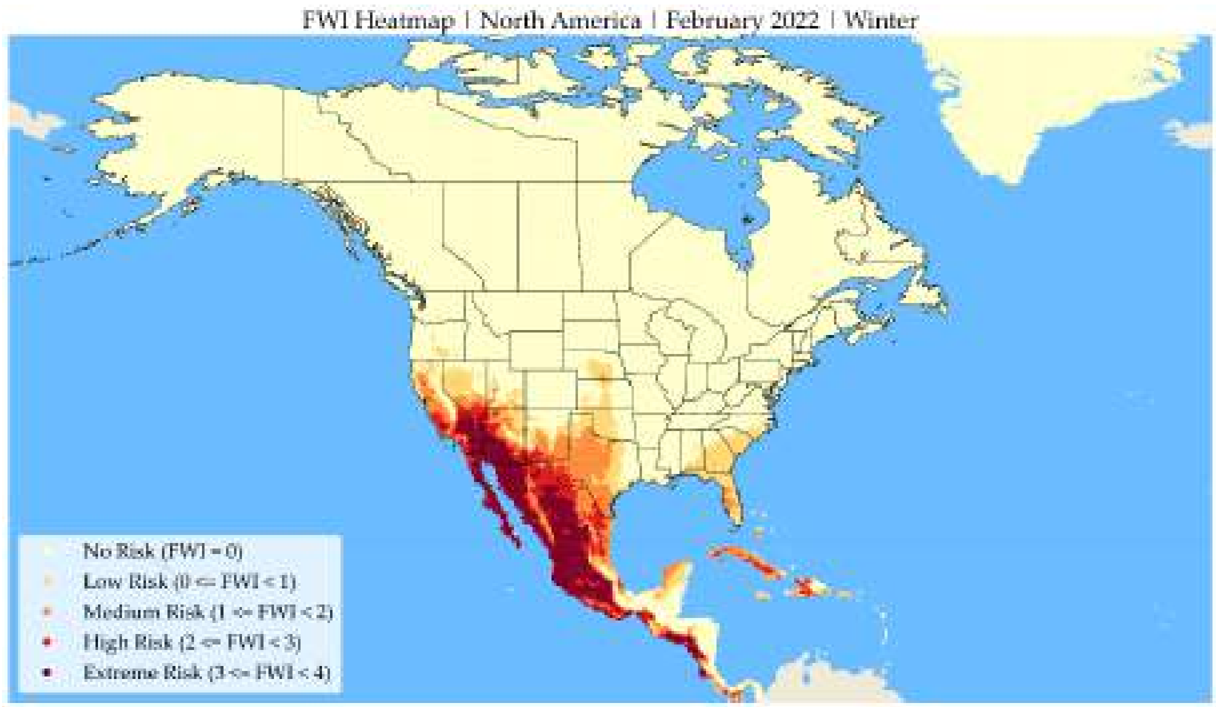

Fig 7: Heatmap of FWI Values for February 2022, in the general Winter season of the region

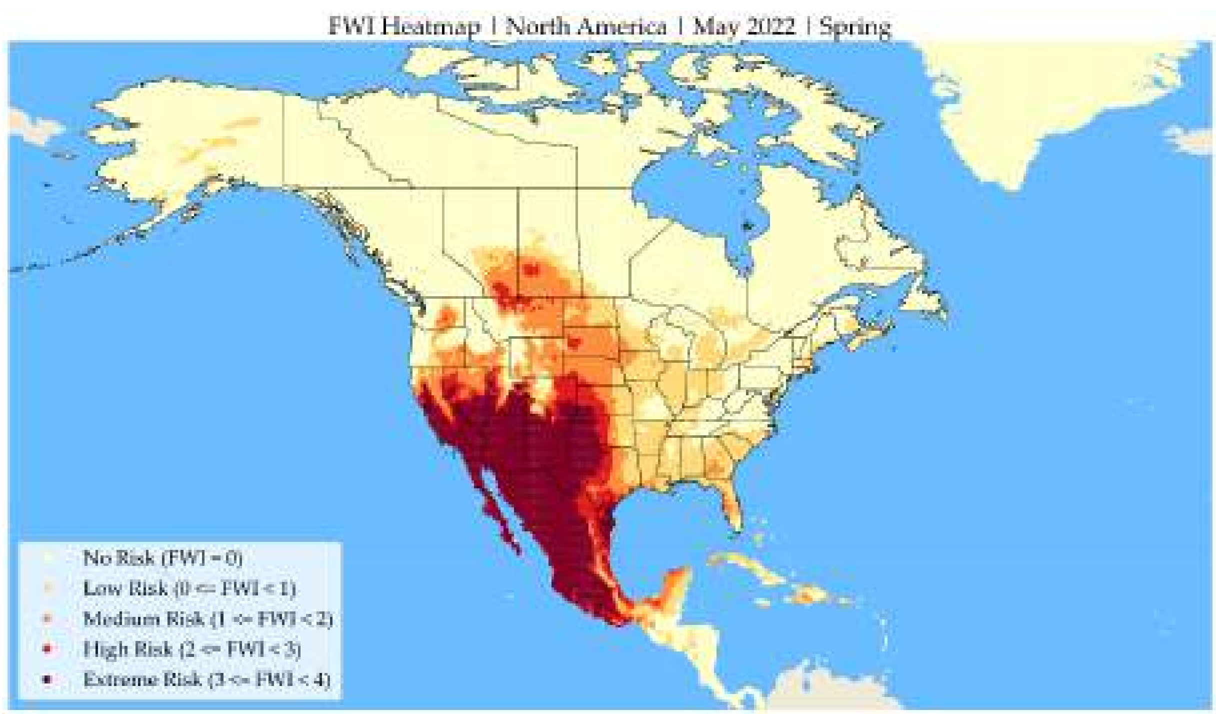

Fig 8: Heatmap of FWI Values for May 2022, in the general Spring season of the region

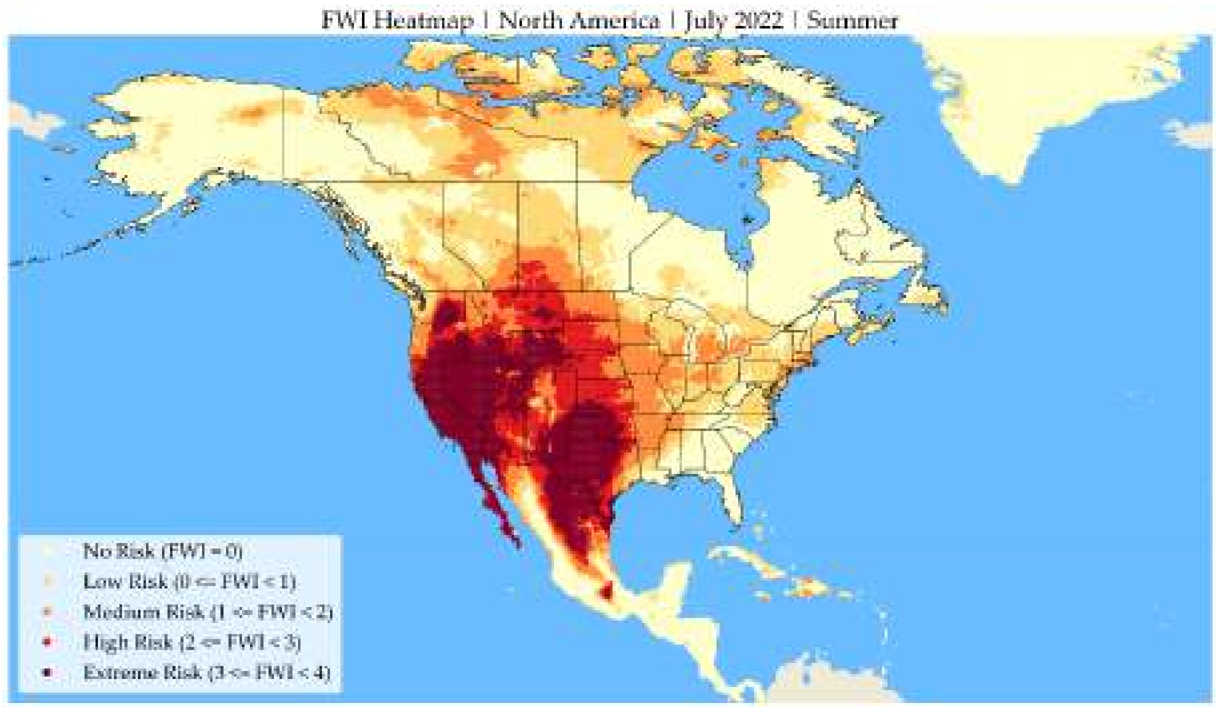

Fig 9: Heatmap of FWI Values for July 2022, in the general Summer season of the region

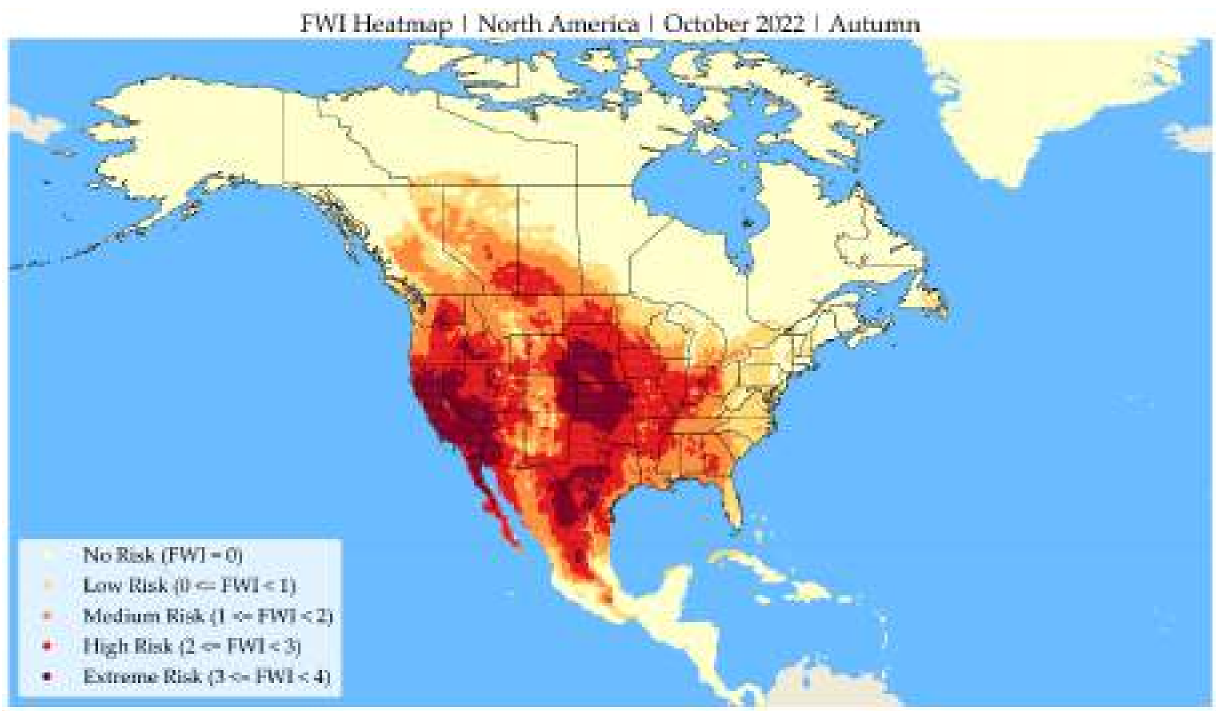

Fig 10: Heatmap of FWI Values for October 2022, in the general Autumn season of the region

In the subsequent tables, each column represents specific information as follows:

1. 'N-Fold Index': This column corresponds to the training fold number utilized in the cross-validation process, which helps assess the model's performance across various subsets of the dataset.

2. 'Month': This column indicates the specific month for which the fire risk predictions are made.

3. 'Year': This column denotes the year for which the fire risk predictions are generated.

4. 'No Risk Percentage Accuracy', 'Low Risk Percentage Accuracy, 'Moderate Risk Percentage Accuracy', 'High Risk Percentage Accuracy', 'Extreme Risk Percentage Accuracy': These columns present the accuracy of the model in predicting the monthly percentage of every risk level within an error margin of 10%; real values are represented in percentages as well.

5. 'Weighted-average Accuracy': This column displays the accuracy of the model for values that fall within an error margin of one risk level.

N-Fold Index	Month	Year	No Risk Percentage Accuracy	Low Risk Percentage Accuracy	Moderate Risk Percentage Accuracy	High Risk Percentage Accuracy	Extreme Risk Percentage Accuracy	Weighted average Accuracy
1	7	2022	52.4	51.9	83.1	92.5	91.7	98.47
2	8	2022	55.5	52.8	81.0	90.4	87.5	95.75
3	9	2022	64.7	63.2	84.2	90.9	89.3	95.45
4	10	2022	69.7	69.6	82.1	86.8	86.6	91.95
5	11	2022	77.6	82.0	89.4	93.0	94.9	95.84
6	12	2022	91.3	90.3	95.6	95.8	95.9	99.20
7	1	2023	92.8	92.9	95.5	95.7	95.6	97.77
8	2	2023	91.9	92.6	96.0	96.8	96.8	99.54
9	3	2023	87.5	91.3	95.0	96.4	95.5	98.61
10	4	2023	81.8	85.1	93.2	95.5	94.9	98.22
11	5	2023	55.3	56.7	82.5	92.6	92.4	97.26
12	6	2023	56.1	52.8	82.1	90.2	88.4	93.86

Table 1: One-Shot Model's Fold Performance for the North America Region

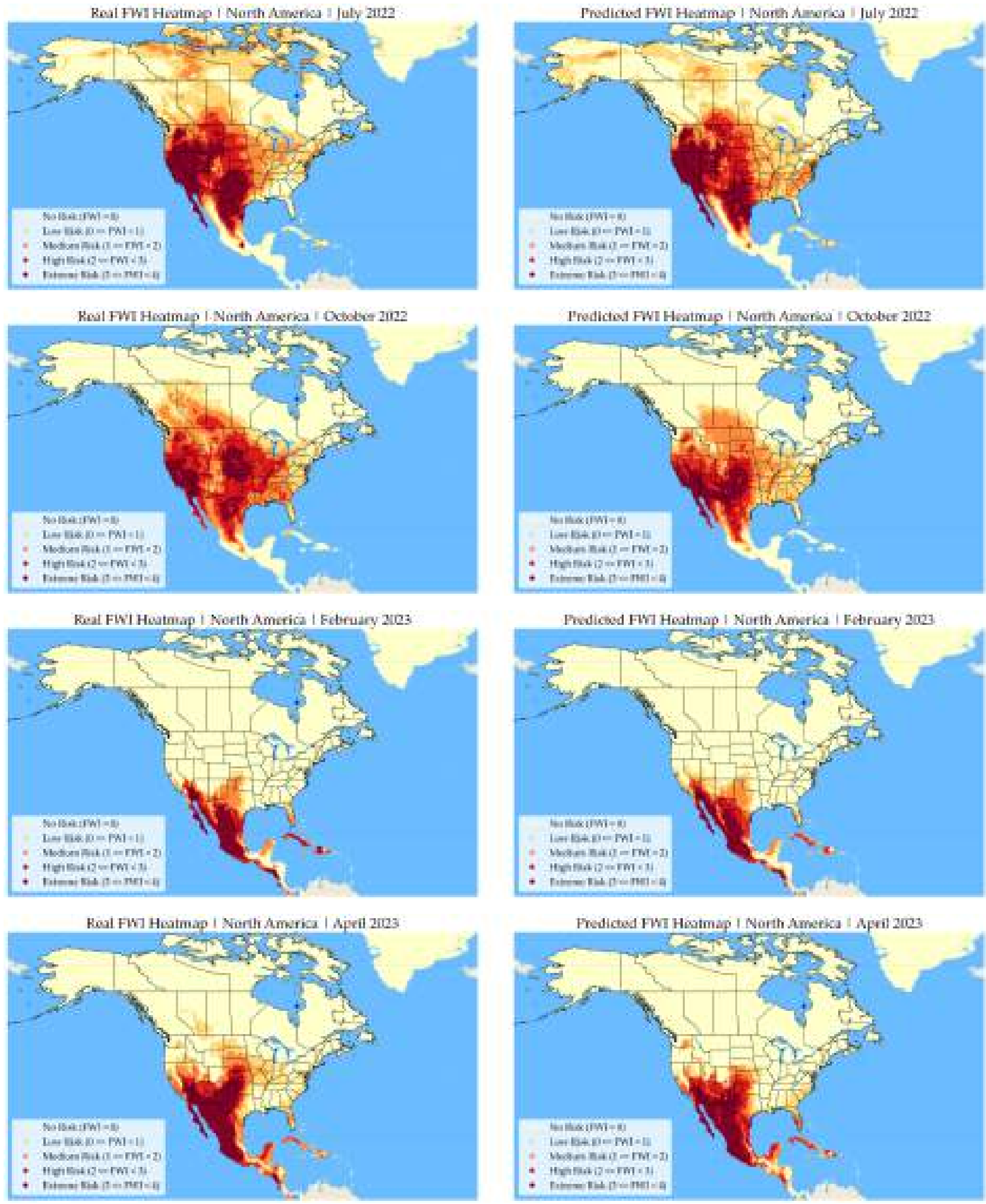

Fig 11-14: Comparative Heatmaps of FWI Values and One-Shot Model Predictions Across Seasons

N-Fold Index	Month	Year	No Risk Percentage Accuracy	Low Risk Percentage Accuracy	Moderate Risk Percentage Accuracy	High Risk Percentage Accuracy	Extreme Risk Percentage Accuracy	Weighted average Accuracy
1	7	2022	50.8	50.5	83.9	92.7	90.4	98.79
2	8	2022	55.2	52.3	82.0	90.6	87.6	96.15
3	9	2022	66.1	65.1	85.0	91.1	89.4	96.65
4	10	2022	71.2	70.2	82.5	87.5	87.0	94.53
5	11	2022	67.7	75.1	85.4	89.2	92.0	91.56
6	12	2022	89.8	88.9	95.3	95.4	95.5	98.82
7	1	2023	93.0	92.6	96.0	96.0	95.7	98.16
8	2	2023	91.3	92.2	96.1	96.8	96.5	99.42
9	3	2023	87.8	91.8	95.0	96.2	95.2	98.57
10	4	2023	80.3	83.6	92.0	95.1	94.0	97.90
11	5	2023	55.3	56.3	82.4	92.9	92.4	96.68
12	6	2023	56.9	52.4	82.8	90.8	88.2	94.72

Table 2: Year-By-Year Model's Fold Performance for the North America Region

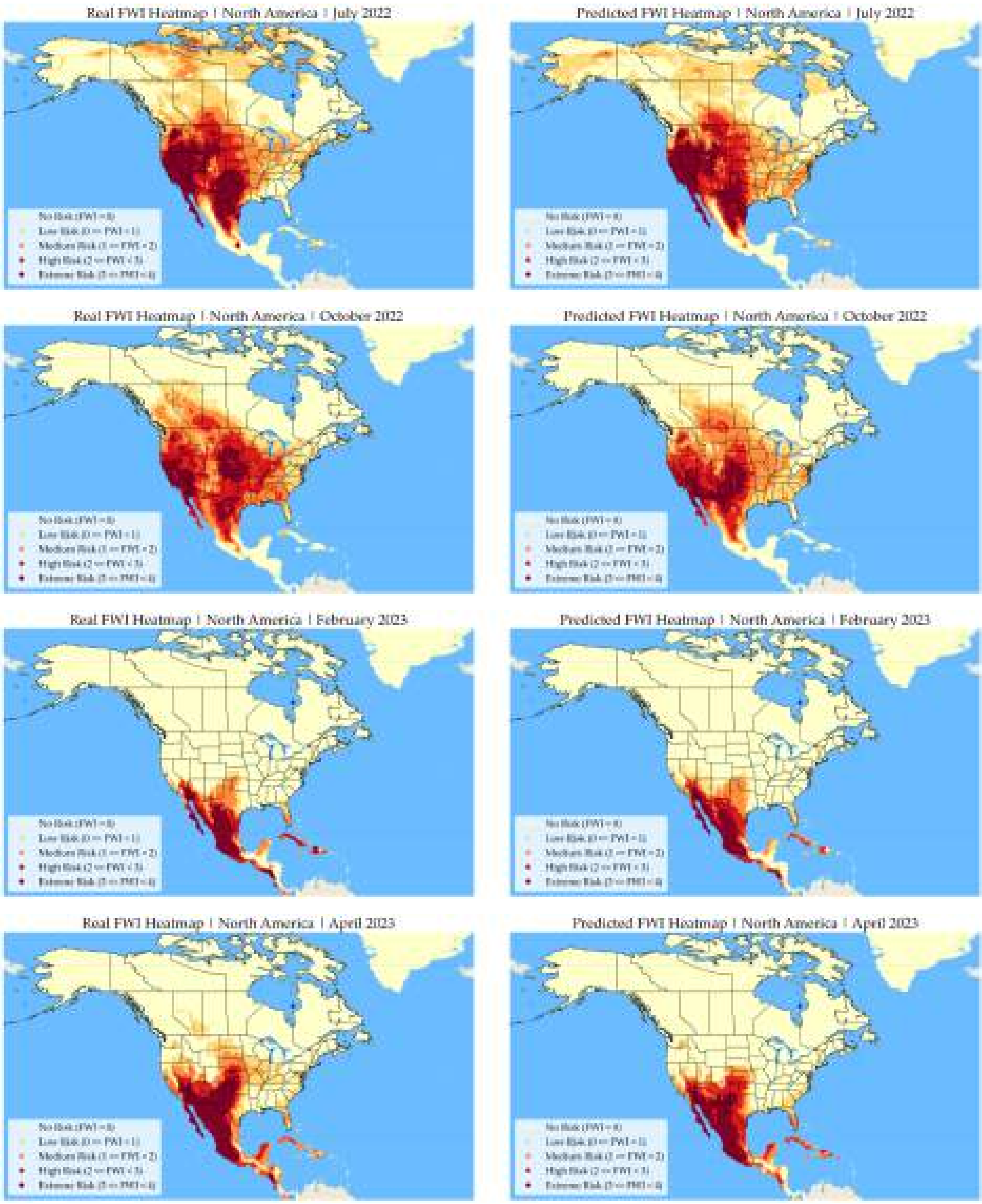

Figure 15-18: Comparative Heatmaps of FWI Values and Year-By-Year Model Predictions Across Seasons

From Table 1 and Table 2, we can immediately notice that our 'Moderate Risk', 'High Risk', and 'Extreme Risk' percentages consistently exhibit accurate trends throughout the folds. However, the accuracies for 'No Risk' and 'Low Risk' percentages vary, following the seasonal patterns of fire occurrences. Higher accuracies align with periods of minimal fire coverage, particularly noticeable during the early months of the year in Figure 1, when fire-prone coverage is low and increases rapidly in later months. Our model accurately distinguishes positive fire risk (High and Extreme) from negative fire risk (No and Low) during winter and spring seasons. Conversely, the model performs well in classifying positive fire risk during months with higher fire-prone coverage but struggles with negative fire risk. In conclusion, both models tend to overestimate fire risk during the Summer and Autumn seasons, while maintaining high accuracy in predicting high and extreme risk coverage throughout all months.

Furthermore, upon analyzing the two models - using One-Shot and Year-By-Year encoding methods - we observe distinct patterns in their importance attribution. The first model heavily emphasizes the 'month_encoded' risk levels for the next month, as expected, since it aims to predict the risk level for the subsequent period. Conversely, it assigns relatively less importance to the current month's 'month_encoded' risk levels and overall risk 'percentage' levels. Notably, the model shows higher importance for the 'No Risk' level in the lagged '(n-11)' variables and 'target_month_encoded' variables, which is mainly attributed to the task at hand for the model: forecasting risk percentages. An interesting to note is that the One-Shot model places a lower emphasis on the 'geo', 'ECO_ID', and 'BIOME_NUM' features, which can be interpreted as the model's ability to generalize over the entire region, regardless of the fuel type and land over, while focusing purely on the historical nature of the data.

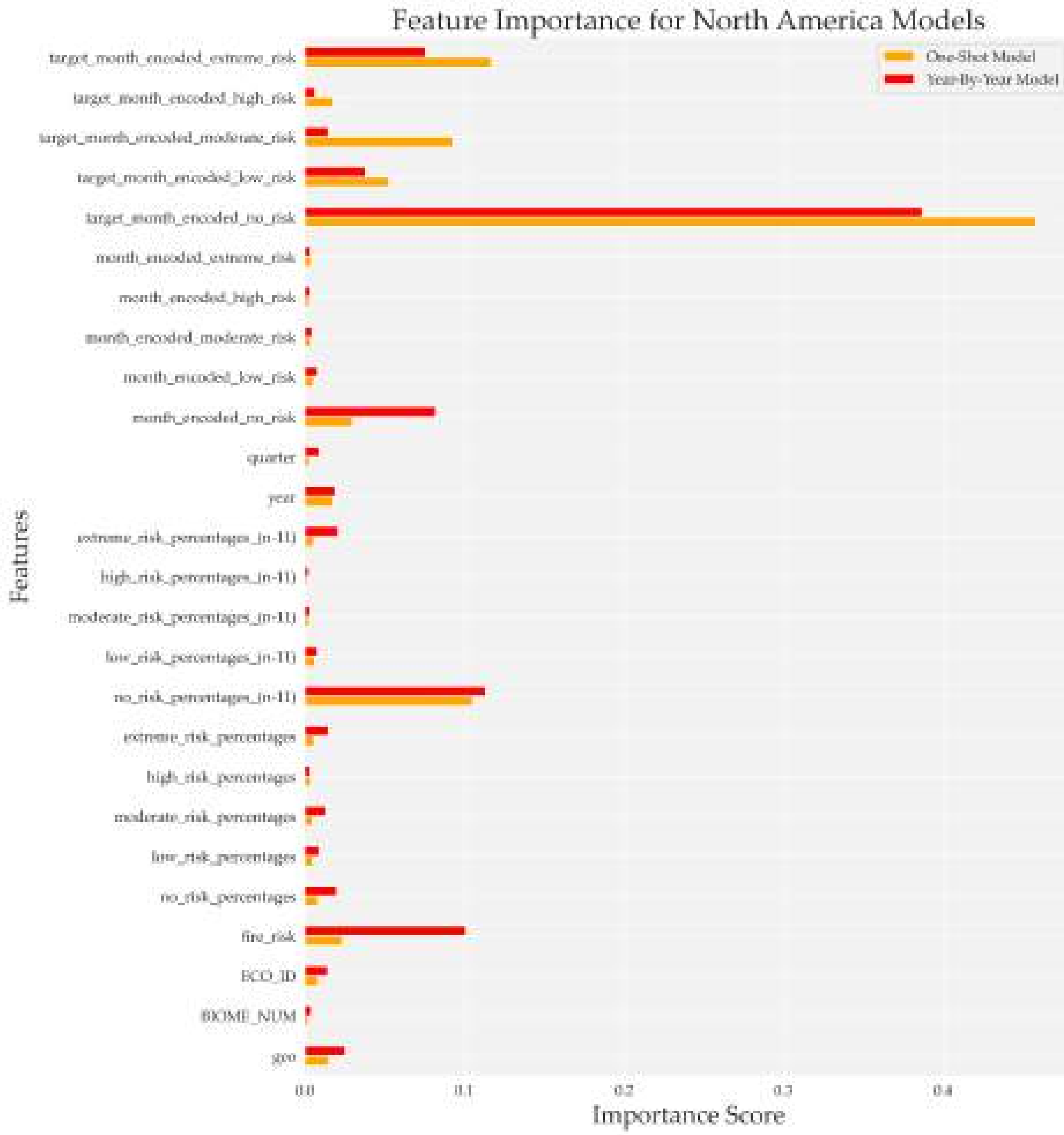

Fig 19: Feature importance for One-Shot and Year-By-Year Encoded forecasting models.

On the other hand, the second model also prioritizes the 'month_encoded' values for the next month, while also placing greater importance on the 'Extreme Risk' encoded values. Interestingly, it attributes a comparatively higher significance to the 'No Risk' and 'Extreme Risk' values for both the current month's 'month_encoded' risk levels and risk 'percentage' levels. This observation could be attributed to the consistent prevalence of 'No Risk' levels across a significant portion of the North American region throughout the entire training data period. As a result, the model tends to assign relatively higher importance to this aspect. But, the important observation here is the ability of the model to prioritize the 'Extreme Risk' level more, since the areas which generally have an extreme fire risk have consistently high 'month_encoded' values for the same. Additionally, the model gives considerable importance to the average risk level, possibly indicating its significance as a critical indicator across various regions.

Overall, both models demonstrate a common focus on predicting the risk level for the next month, while accounting for the unique characteristics of the North American Region. These observed differences and similarities in importance attribution provide valuable insights into the underlying decision-making processes of the models and offer a deeper understanding of their performance in the context of wildfire risk prediction.

Upon closer inspection of the One-Shot and Year-By-Year Models, we can see that there is minimal difference in the weighted accuracy tables. This is because our accuracy metrics only capture how well the models are able to differentiate between Positive and Negative risk levels. However, when we analyze the comparison plots of the N-Folds, we can see that the Year-By-Year model predicts the exact risk class level with greater precision than the One-Shot model. We can assign this capability to the latter model being able to capture the random, noisy change in the fire risk across the fire seasons. These comparisons are made more clearly in the following table that places the accuracies of both models side-by-side across the N-Folds.

N-Fold Index	Month	Year	One-Shot Accuracy	Year-By-Year Accuracy	Accuracy Comparison (One-Shot - Year-By-Year)
1	7	2022	98.47	98.79	-0.32
2	8	2022	95.75	96.15	-0.40
3	9	2022	95.45	96.65	-1.20
4	10	2022	91.95	94.53	-2.58
5	11	2022	95.84	91.56	4.28
6	12	2022	99.20	98.82	0.37
7	1	2023	97.77	98.16	-0.39
8	2	2023	99.54	99.42	0.12
9	3	2023	98.61	98.57	0.04
10	4	2023	98.22	97.90	0.33
11	5	2023	97.26	96.68	0.59
12	6	2023	93.86	94.72	-0.86

Table 3: Weighted-average accuracy comparison between One-Shot and Year-By-Year models across N-Folds

In order to showcase the accuracy and performance of our fire risk prediction models, as a whole, we consider a real-life example of how these models would be used. First, we trained the model using data up to May 2023 only. Subsequently, we put the model to the test by predicting fire risk values for the next two months, June and July, of the same year. By comparing these predictions with the actual fire risk values observed to date, we have generated the following accuracy comparison plots and tables.

	One-Shot Model	Year-By-Year Model
Month	7	7
Year	2023	2023
No Risk Percentage Accuracy	54.07	50.29
Low Risk Percentage Accuracy	51.93	48.33
Moderate Risk Percentage Accuracy	81.27	77.63
High Risk Percentage Accuracy	91.62	89.95
Extreme Risk Percentage Accuracy	90.79	88.35
Weighted-average Accuracy	98.15	96.85

Table 4: Percentage and Weighted-average Accuracy comparison between One-Shot and Year-By-Year models for July 2023

Our test comparing the prediction of real-time July 2023 data with both the One-Shot and Year-By-Year models revealed compelling insights. The One-Shot model outperformed the Year-By-Year model across all aspects, demonstrating its superior predictive capabilities. In the feature importances analysis of both models, we observed that the One-Shot model exhibited better generalization over the entire North American region. Unlike the Year-By-Year model, the One-Shot model showed reduced significance on the previous month's fire risk values. This finding suggests that the One-Shot model can capture and predict fire risk more holistically, considering broader geographical trends, and highlights the significance of considering long-term temporal dynamics in fire risk assessment, as it enables the model to better generalize and adapt to varying conditions over time.

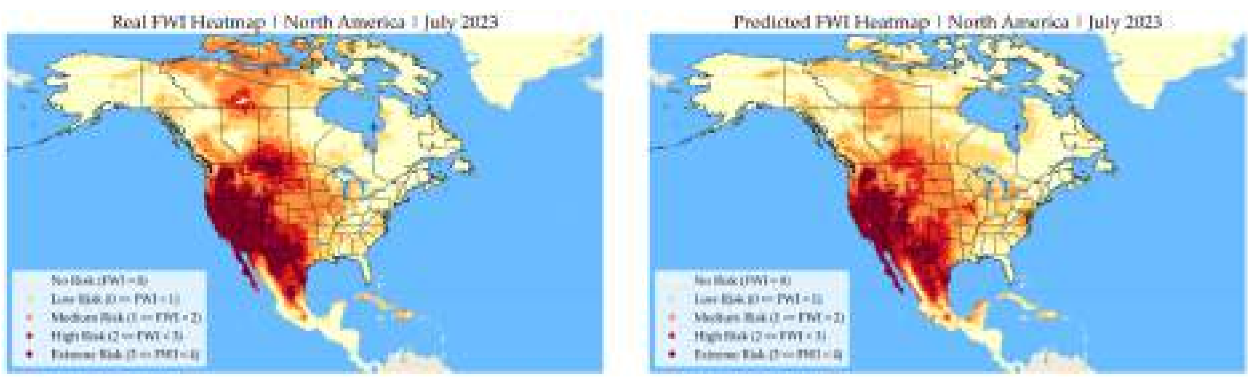

Figure 20: Comparative Heatmaps of FWI Values and One-Shot Forecasts for July 2023

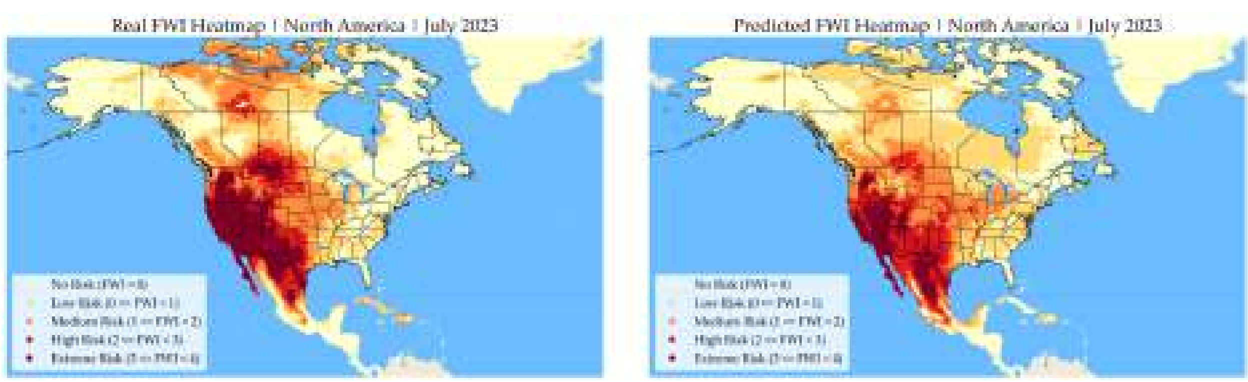

Figure 21: Comparative Heatmaps of FWI Values and Year-By-Year Forecasts for July 2023

Having established the accuracy and validation of our model through a comprehensive analysis of the training process folds, we now turn our attention to comparing our model forecasts with widely used government agency charts.

In this segment, we examine how our model's fire risk predictions align with the forecasts generated by reputable agencies such as the Canadian Wildland Fire Information System (CWFIS) and National Interagency Fire Center (NIFC). By juxtaposing these results, we aim to demonstrate the practical applicability and reliability of our model in predicting fire risk on a larger scale and under real-world conditions. First, we will present a comparative analysis between our model's fire risk forecasts and the Canadian Seasonal to Inter-annual Prediction System (CanSIPS) plots. Although these forecasts are generated on different scales, their overall interpretations of fire risk exhibit notable similarities.

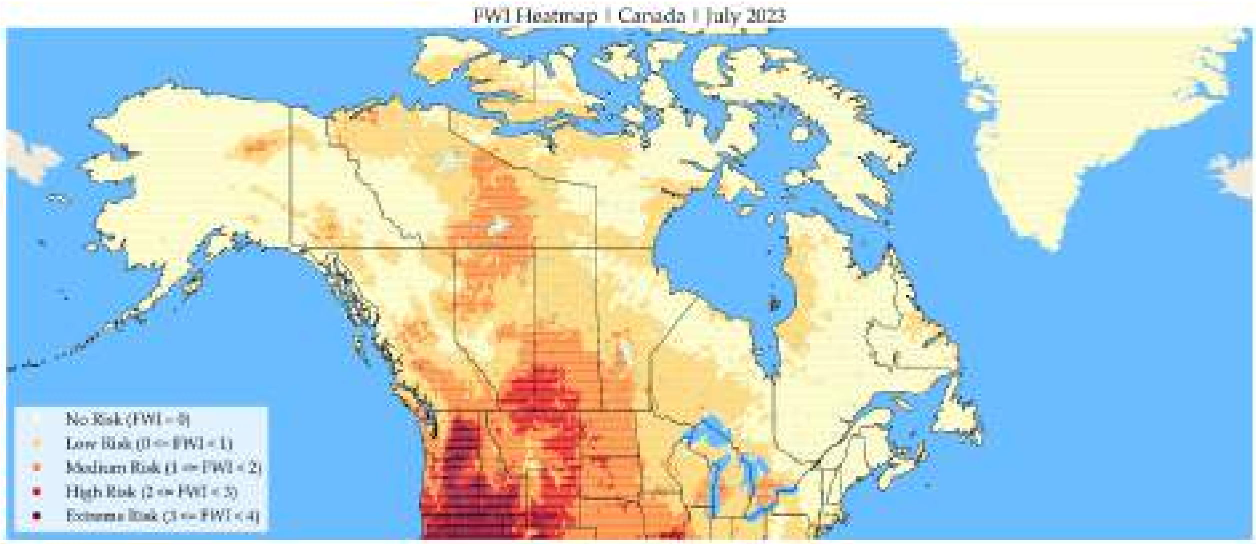

Fig 22: Heatmap of FWI values forecast across Canada for July 2023

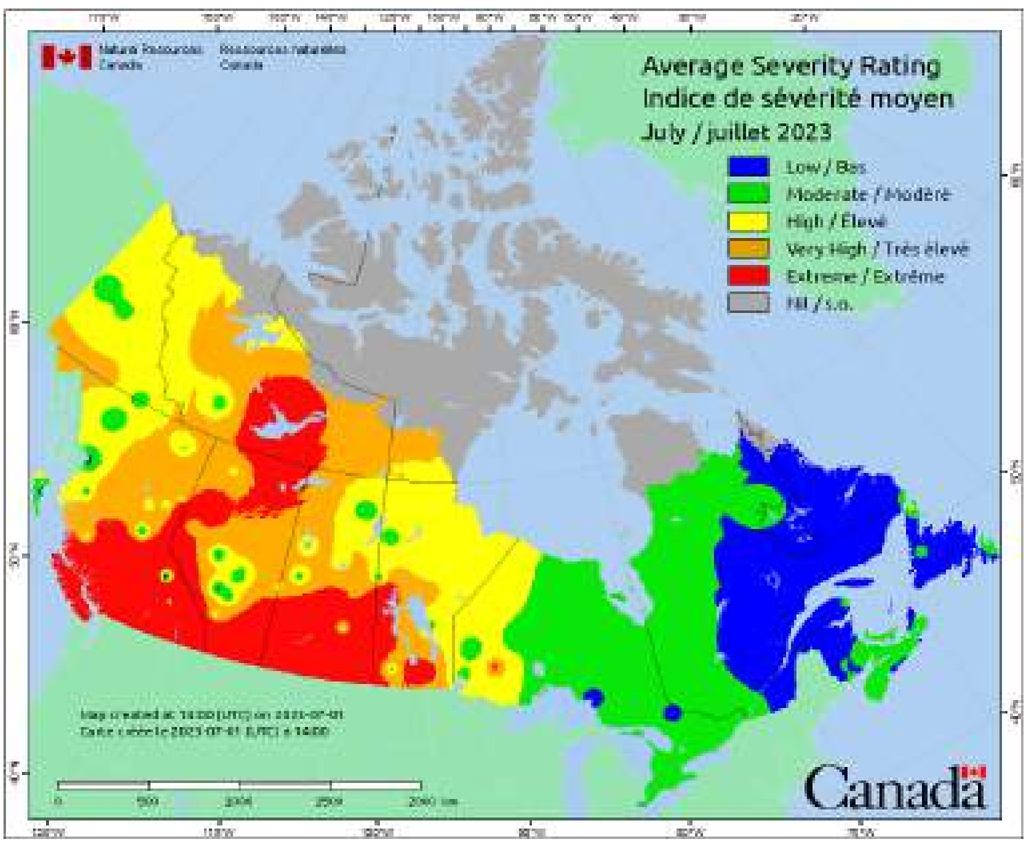

Fig 23: Heatmap of FWI values forecast across Canada (CWFIS) for July 2023

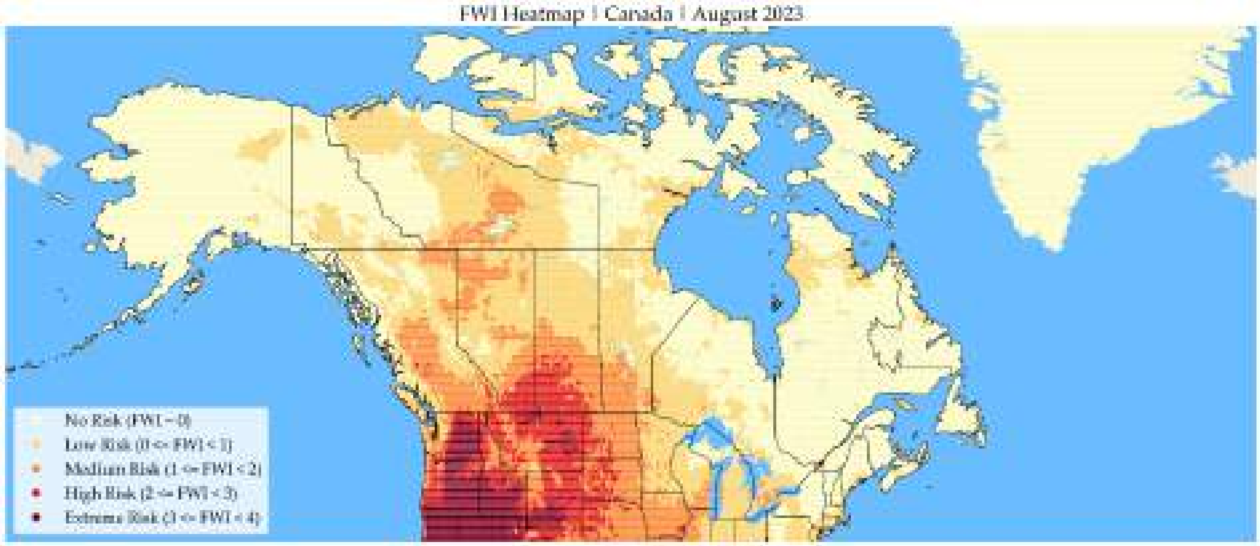

Fig 24: Heatmap of FWI values forecast across Canada for August 2023

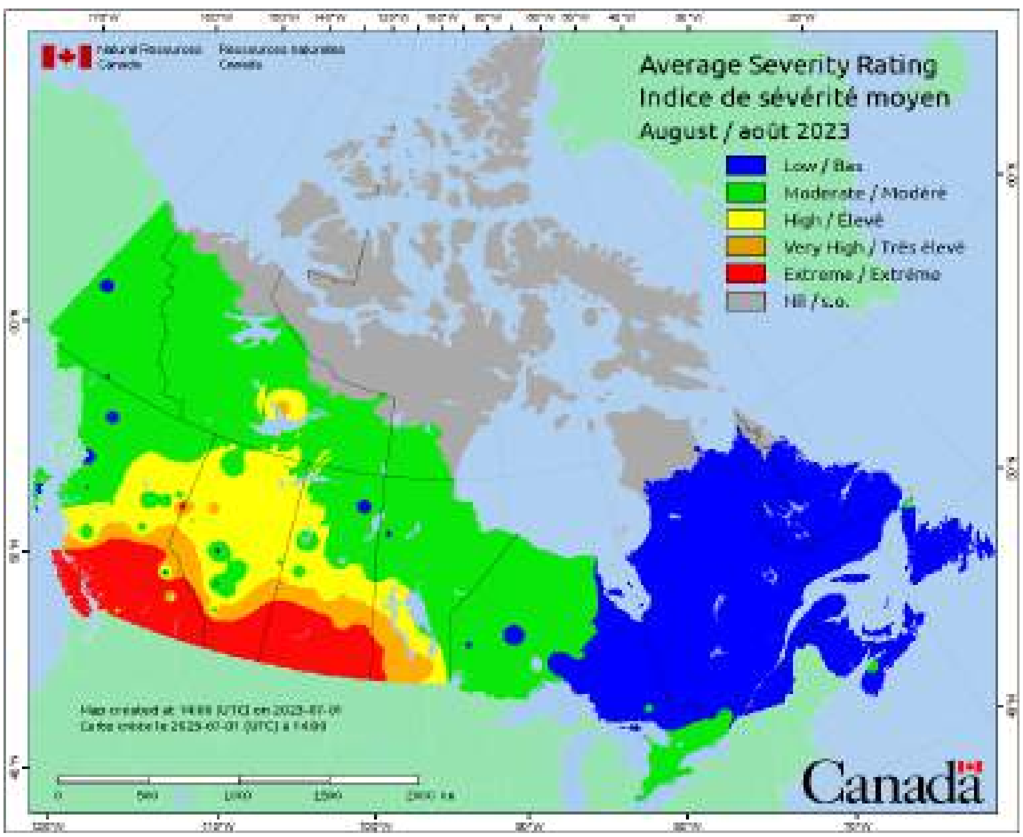

Fig 25: Heatmap of FWI values forecast across Canada (CWFIS) for August 2023

For the July forecast comparisons, both models identify an area of increased risk around the Great Salt Lake in the Northwest Territories. Additionally, both forecasts indicate elevated fire risk areas across the provinces of British Columbia, Alberta, and Saskatchewan. Ontario and Quebec show comparable regions of No Risk to Low Risk. In Yukon and Northwest Territories, similar-shaped spaces of Moderate to High risk are identified.

Coming to the August forecast comparisons, we can clearly see identical High Risk to Extreme Risk areas covering British Columbia, Alberta, and Saskatchewan. The upper parts of the same states have a comparatively lower risk but still show the presence of fire occurrence in both forecasts. Here, we see Quebec having almost No Risk throughout August, as is predicted by our models as well. We can also see hotspots of increased risk around the Great and Lesser Salt Lakes that match the predictions made by the Canada forecast models.

Notably, the Canada forecasts are generated by an ensemble of weather prediction models by the Canadian Centre for Climate Modelling and Analysis (CCCMA) and the Canadian Meteorological Centre (CMC). In contrast, our models solely rely on historical FWI data. Despite this distinction, our models maintain a high level of accuracy in comparison to the weather-based models, reinforcing the robustness of our approach.

Overall, the alignment between the two sets of forecasts, coupled with our model's ability to demonstrate good validation accuracy with the real Canada maps, underscores the effectiveness and reliability of our fire risk prediction model. And our models which purely leverage historical FWI data are able to maintain high accuracy with the weather based models too.

Having thoroughly validated our model's fire risk forecasts with the CanSIPS charts, we now turn our focus to evaluating our model's performance for the entire North American region. In this segment, we will compare our model's predictions with the widely recognized fire risk charts provided by the National Interagency Fire Center (NIFC) for the North American region.

In contrast to our forecasting models, which provide absolute fire risk values where greater redness indicates higher risk, the NIFC offers 'relative' fire outlooks. "The outlook identifies areas by month for the next four months with above, below, and near normal significant fire potential" (National Interagency Fire Center, n.d.). In their maps, areas marked in red signify greater risk than normal, while green areas indicate lesser risk than normal. On the other hand, our models offer precise numerical risk values, providing a detailed evaluation of fire risk levels but may require additional context to gauge risk severity in comparison to typical conditions.

The forecasted heatmaps maps for North America for July, August, and September months of the year 2023 are compared with their corresponding official NIFC forecasts as shown below.

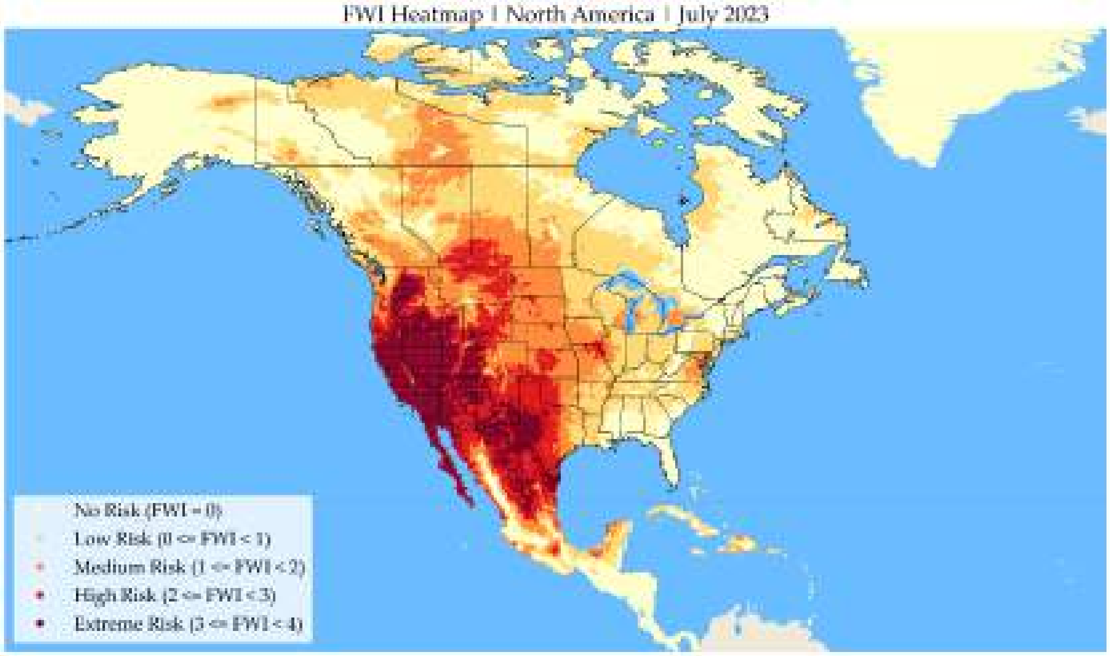

Fig 26: Heatmap of FWI values forecast across North America for July 2023

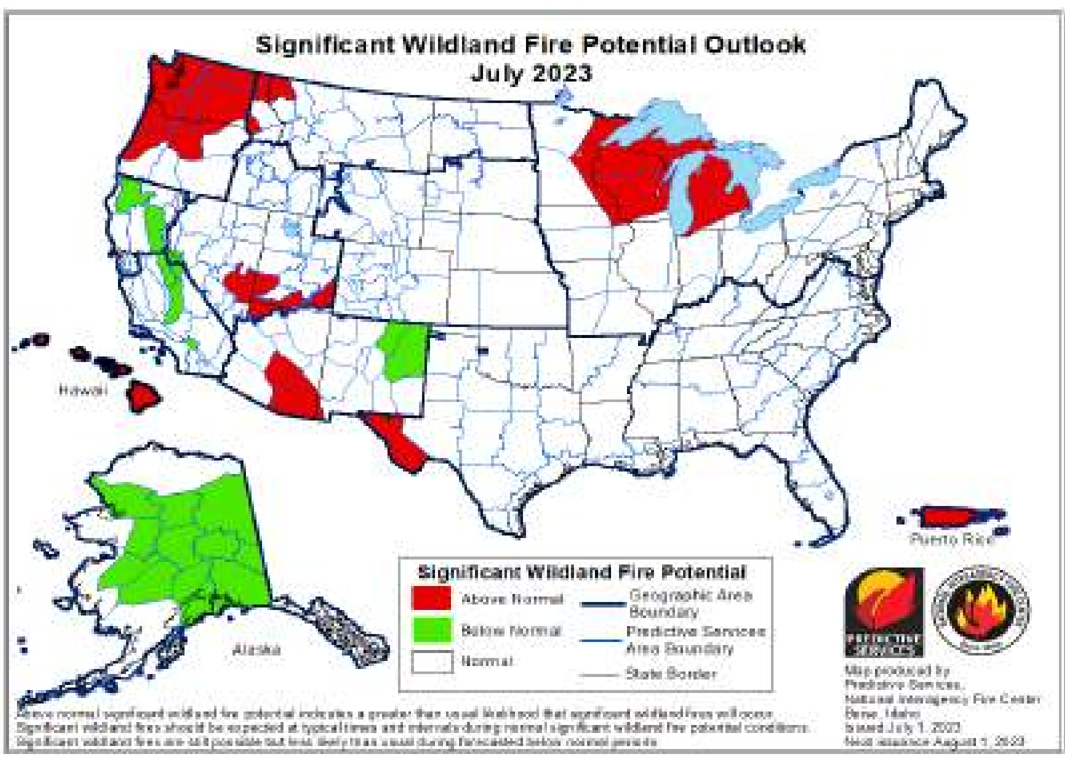

Fig 27: NIFC Fire Potential Outlook across United States of America for July 2023

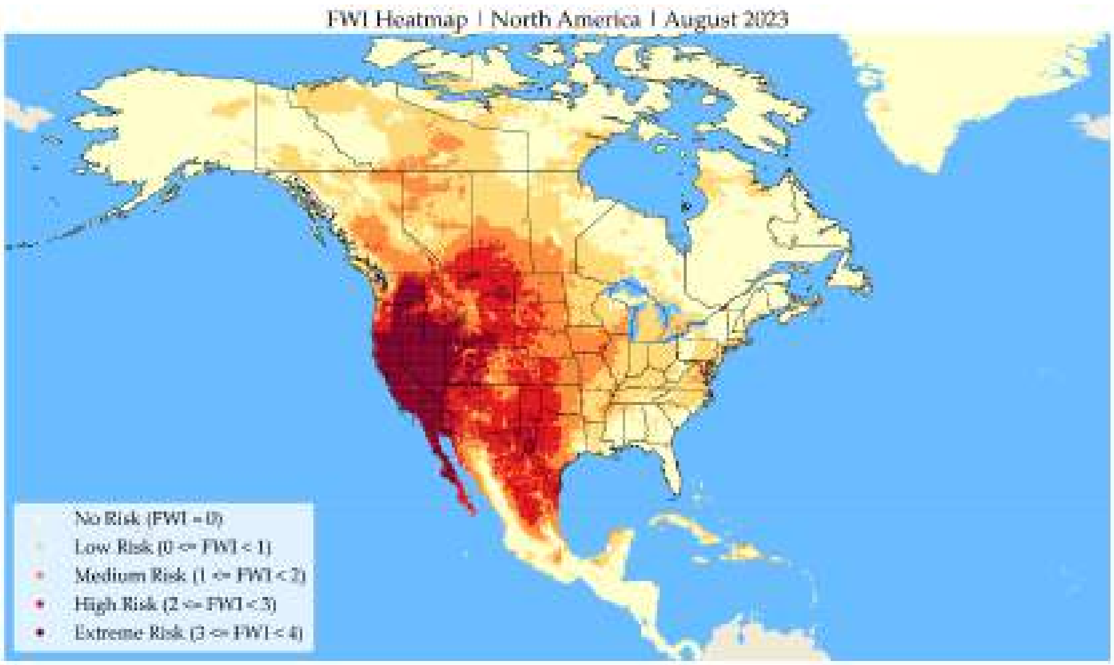

Fig 28: Heatmap of FWI values forecast across North America for August 2023

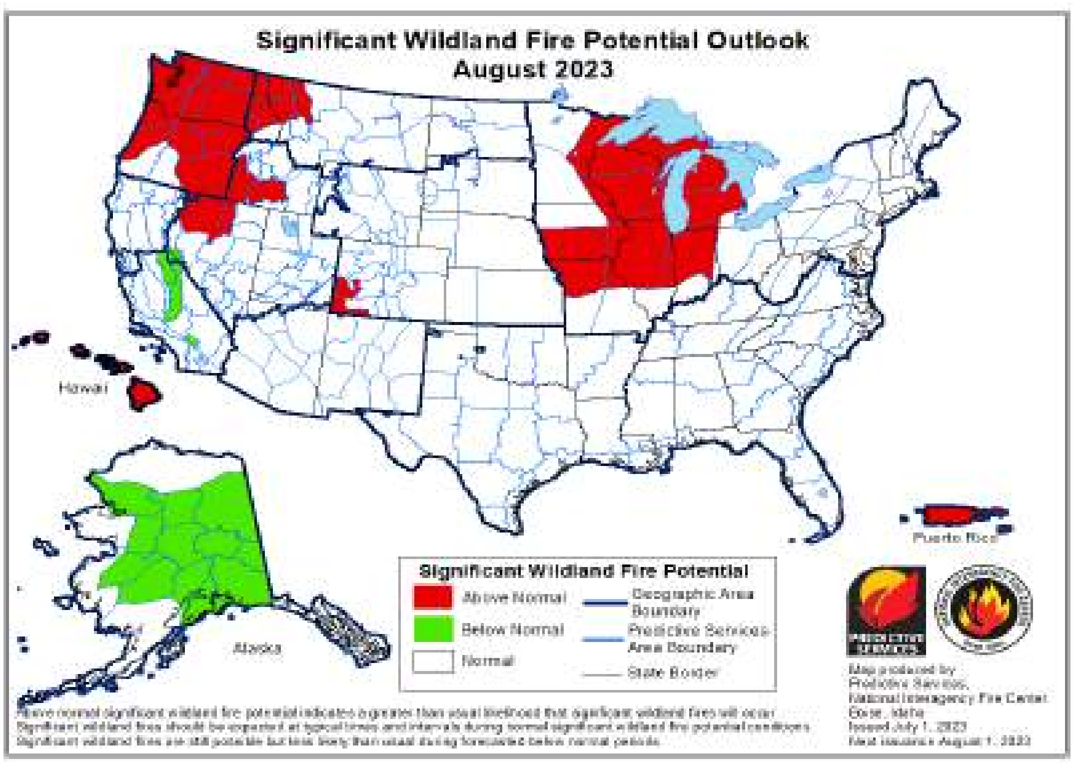

Fig 29: NIFC Fire Potential Outlook across the United States of America for August 2023

Based on our comparison between the fire risk forecasts generated by our model and the National Interagency Fire Center (NIFC) agency's forecast map for the United States, we have made several key observations. In July, the forecasted fire risk due to weather aggravates in regions such as Washington, Oregon, Arizona, and parts of Texas, aligning with the summer season for North-West, South-West, and North America. Conversely, we found that the fire risk reduces in Arizona and some parts of Texas in August. Notably, the central region of California exhibits lower risk compared to its surrounding areas, a trend also evident in our model's plots.

To gain deeper insights, we compared our plots for July and August, focusing on areas of significant change and cross-referencing with the NIFC August relative fire outlook. We observed that in Washington and Oregon, there is a greater extreme risk outlook from July to August, corroborated by the NIFC August plot. Similarly, the extent of the green area in California decreases from July to August, reflecting the findings in our plots. Noteworthy regions with greater fire outlook include parts of Utah and Arizona. In New Mexico, our model places the region at moderate risk for both months, consistent with the relative NIFC plots.

These consistent and validated observations reinforce the high accuracy and reliability of our model's fire risk predictions when compared to the widely recognized NIFC agency's forecast maps for the United States. The alignment between our model's forecasts and the NIFC predictions demonstrates the practical applicability and potential impact of our model in assisting wildfire risk assessment and management efforts across the USA.

Conclusion

In conclusion, our research paper has shed light on the critical aspect of temporal information in fire risk prediction models. By developing and analyzing two distinct models, namely One-Shot and Year-By-Year encoding methods, we have uncovered significant implications for accurately predicting fire risk. Our findings have demonstrated that the temporal aspect plays a crucial role in enhancing predictive performance, with the One-Shot model exhibiting superior accuracies and more favorable plots in our comparisons. Moreover, an equally vital aspect of our investigation lies in understanding the differences in importance attribution provided by the models. These important disparities reveal how various encoding methods influence the model training process, providing valuable insights into their decision-making mechanisms. Such observations are essential for advancing the field of fire risk prediction and paving the way for more effective and informed model design. In essence, this research paper underscores the significance of leveraging temporal information and carefully selecting encoding methods in the development of robust fire risk prediction models. By harnessing the temporal dimension and embracing innovative approaches, we can enhance the accuracy and reliability of these models, contributing to more effective wildfire management and protection measures in the future. In addition to the insights and achievements presented in this research paper, there are several promising avenues for future development and improvement in the field of fire risk prediction. The next steps forward involve integrating forecasted weather data, along with crucial factors such as drought, lightning, and extreme weather events, to create a more comprehensive and robust predictive model. Incorporating forecasted weather data into the existing models would allow us to account for the dynamic nature of wildfire risk, capturing how weather conditions may influence fire behavior over time. By integrating historical trends and forecasted weather information, the predictive capabilities of our models can be further refined, providing more accurate and timely fire risk assessments. Moreover, the inclusion of additional critical features, such as drought indices and lightning data, can significantly enhance the model's predictive power. Drought indices can offer valuable insights into the long-term moisture conditions, which play a vital role in determining the susceptibility of an area to wildfire. Similarly, considering lightning data will enable us to assess the potential ignition sources and their contribution to fire risk. Our work serves as a stepping stone towards a deeper understanding of fire risk prediction, encouraging further exploration and advancements in this vital domain. These future developments have the potential to revolutionize wildfire management strategies, enabling proactive and adaptive approaches to minimize the impact of wildfires on our communities and natural ecosystems.

References

1. Reuters. (2021, August 11). Algeria wildfire toll rises to 65, state television says. https://www.reuters.com/world/africa/algeria-wildfire-toll-rises-65-state-television-says-2021-08-11/

2. Statista. (n.d.). Area of acres burnt due to wildland fires in the US. https://www.statista.com/statistics/203990/area-of-acres-burnt-due-to-wildland-fires-in-the-us/

3. Jaque Castillo, E., Fernández, A., Fuentes Robles, R., and Ojeda, C. G.: Data-based wildfire risk model for Mediterranean ecosystems – case study of the Concepción metropolitan area in central Chile, Nat. Hazards Earth Syst. Sci., 21, 3663–3678, https://doi.org/10.5194/nhess-21-3663-2021

4. Sihan Li et al 2021 Environ. Res. Lett. 16 094051 https://iopscience.iop.org/article/10.1088/1748-9326/ac1e3a

5. Harris S, Lucas C. Understanding the variability of Australian fire weather between 1973 and 2017. PLoS One. 2019 Sep 19;14(9):e0222328. https://doi.org/10.1371/journal.pone.0222328

6. Chen Y, Randerson JT, Morton DC, DeFries RS, Collatz GJ, Kasibhatla PS, Giglio L, Jin Y, Marlier ME. Forecasting fire season severity in South America using sea surface temperature anomalies. Science. 2011 Nov 11;334(6057):787-91. https://doi.org/10.1126/science.1209472.

7. Alan Mandal, Grzegorz Nykiel, Tomasz Strzyzewski, Adam Kochanski, Weronika Wrońska, Marta Gruszczynska, and Mariusz Figurski.

"High-resolution fire danger forecast for Poland based on the Weather Research and Forecasting Model" International Journal of Wildland Fire (2022): 149-162. https://doi.org/10.1071/WF21106

8. Tian X.-R., Zong X.-Z., Shu L.-F., Wang M.-Y., & Zhao F.-J. (2020). Impacts of ENSO events on forest fire weather of China. Ying yong sheng tai xue bao [The journal of applied ecology], 31(5), 1487–1495. https://doi.org/10.13287/j.1001-9332.202005.015

9. Bedia, J., Herrera, S., Martín, D.S. et al. Robust projections of Fire Weather Index in the Mediterranean using statistical downscaling. Climatic Change 120, 229–247 (2013). https://doi.org/10.1007/s10584-013-0787-3

10. Calheiros T, Pereira MG, Nunes JP. Assessing impacts of future climate change on extreme fire weather and pyro-regions in Iberian Peninsula. Science of The Total Environment. 2021 Feb 1;754:142233. http://dx.doi.org/10.1016/j.scitotenv.2020.142233

11. Hantson, S., Arneth, A., Harrison, S. P., Kelley, D. I., Prentice, I. C., Rabin, S. S., Archibald, S., Mouillot, F., Arnold, S. R., Artaxo, P., Bachelet, D., Ciais, P., Forrest, M., Friedlingstein, P., Hickler, T., Kaplan, J. O., Kloster, S., Knorr, W., Lasslop, G., Li, F., Mangeon, S., Melton, J. R., Meyn, A., Sitch, S., Spessa, A., van der Werf, G. R., Voulgarakis, A., and Yue, C.: The status and challenge of global fire modelling, Biogeosciences, 13, 3359–3375, https://doi.org/10.5194/bg-13-3359-2016

12. Miller, J., Böhnisch, A., Ludwig, R., and Brunner, M. I.: Climate change impacts on regional fire weather in heterogeneous landscapes of Central Europe, Nat. Hazards Earth Syst. Sci. Discuss. [preprint], https://doi.org/10.5194/nhess-2023-51

13. Fernandes, K., Bell, M., & Muñoz, Á. G. (2022). Combining precipitation forecasts and vegetation health to predict fire risk at subseasonal timescale in the Amazon. In Environmental Research Letters (Vol. 17, Issue 7, p. 074009). IOP Publishing. https://doi.org/10.1088/1748-9326/ac76d8

14. Richardson, D., Black, A. S., Irving, D., Matear, R. J., Monselesan, D. P., Risbey, J. S., Squire, D. T., & Tozer, C. R. (2022). Global increase in wildfire potential from compound fire weather and drought. Npj Climate and Atmospheric Science, 5(1), 1-12. https://doi.org/10.1038/s41612-022-00248-4

15. Lagouvardos, K., Kotroni, V., Giannaros, T. M., & Dafis, S. (2019). Meteorological Conditions Conducive to the Rapid Spread of the Deadly Wildfire in Eastern Attica, Greece. In Bulletin of the American Meteorological Society (Vol. 100, Issue 11, pp. 2137–2145). American Meteorological Society. https://doi.org/10.1175/bams-d-18-0231.1

16. San-Miguel-Ayanz, J., Houston-Durrant, T., Boca, R., Maianti, P., Libertà, G., Oom, D., Branco, A., de Rigo, D., Ferrari, D., Roglia, E., Scionti, N., 2023. Advance report on forest fires in Europe, Middle East and North Africa 2022. Publications Office of the European Union, Luxembourg. ISBN 978-92-68-02143-9, https://doi.org/10.2760/091540

17. *Fire climate regions*. NWCG. (n.d.). https://www.nwcg.gov/publications/pms425-1/fire-climate-regions

18. Natural Resources Canada https://cwfis.cfs.nrcan.gc.ca/maps/forecasts

19. Title: Map of the Provinces and Territories of Canada. (n.d.). In Wikipedia, The Free Encyclopedia. Retrieved July 29, 2023, from https://en.wikipedia.org/wiki/Provinces_and_territories_of_Canada

20. Title: Map of the States and Territories of the United States of America. (n.d.). In Wikipedia, The Free Encyclopedia. Retrieved July 29, 2023, from https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States

21. Google Earth Engine. (n.d.). RESOLVE_ECOREGIONS_2017. Retrieved from https://developers.google.com/earth-engine/datasets/catalog/RESOLVE_ECOREGIONS_2017

22. Los Angeles Times. (2019, January 12). *California's Camp Fire was the costliest global disaster last year, Insurance Report shows*. Los Angeles Times. https://www.latimes.com/local/lanow/la-me-ln-camp-fire-insured-losses-20190111-story.html

23. Paradise Home Page | Paradise, CA. (n.d.). https://www.townofparadise.com/sites/default/files/fileattachments/community/page/22321/townofparadise_campfire_afteractionreport_101920.pdf