Assessing the Impact of Human Inhibition Factors on Wildfire Risk Prediction using Deep Learning

Anant Gupta[1,2]

[1] William A. Shine Great Neck South High School

[2]Carnegie Mellon University

Acknowledgments

Special thanks to Nicole Spinelli for all her comments and suggestions for my manuscript and Professor Shixiang Zhu for his immeasurable support and mentorship

Personal Section

Driving through Canada, we spotted a wildfire burning along the highway. Firefighters arrived quickly and had it out within five minutes. That night in the hotel, I turned on the TV and the first thing I saw was a news report about California's wildfire crisis. From these two events, I grew curious and began voraciously consuming literature on the topic. For my high school's research program, we usually end up picking projects over the summer and for me, this series of events combined with my findings from my initial literature review led me to researching how wildfire risk is predicted and how its predictions can be improved. I performed my research over the summer mostly virtually, with guidance from a professor from Carnegie Mellon (who I reached out to through cold-calling) and in-person in the first quarter of the school year.

Most of my math and science learning came from pursuing research problems that interested me. Whether it was path-planning algorithms in 11th grade, front-end development in 10th grade, or machine learning in 12th grade, I learned by diving deep into literature and learning from tutorials to apply what I learned rather than learning abstract concepts. These experiences of actually seeing math and science used practically, whether in robotics or a research internship, proved fundamental to how my passion for math and science developed.

To my fellow high school researchers: focus on the work itself, not the awards. In fact, let completing your project be the reward. Take pride in what you create and give your soul and best effort to what interests you, even without guaranteed recognition. Do the extra step, learn more than what the textbook tells you, apply what you have learned in some way. And, most of all, remember that the way you approach your research often reflects how you'll approach everything else in life.

Abstract:

The increasing frequency and intensity of wildfires in California have led to severe economic, environmental, and public health damage. To help governments reduce the negative consequences of wildfires, improving wildfire risk prediction is imperative. While currently-used wildfire prediction models have shown potential in predicting wildfires, they typically do not encompass human inhibition factors (number of firefighting stations, acres of land treated by fuel removal strategies, etc.). Thus, this study sought to improve predictive accuracy through the novel inclusion of these human inhibition factors. From this, an LSTM model was created for wildfire risk prediction. A RF model was used as a comparison benchmark, and to observe if the impact of human inhibition factors differed with varying model architectures. As a follow up, this study utilized SHAP analysis and mean decrease in impurity (MDI) both with and without human inhibition factors to evaluate which variables were the strongest predictors of wildfire risk. These models were trained with meteorological, terrain, environmental and anthropogenic variables sourced from publicly available datasets regarding California over the time period of 2019 to 2022. These models were then used to predict where and when wildfires occurred in the year 2023 with a spatial resolution of 76 tiles and a temporal resolution of 4 months. These predictions were cross-validated based on the number of wildfires that actually occurred in 2023. With an F1-score of 0.85, AUROC of 0.90, AUPR of 0.85, a robust LSTM model was developed that can accurately forecast the likelihood of wildfires occurring in 76 tiles of California over a 4 month time scale.

Incorporating human inhibition factors for the LSTM model greatly reduced false negatives, improving F1-score by 0.07 relative to the absence of human inhibition factors; in comparison, the RF model became more conservative in its predictions with increased false negatives and reduced false positives, leading to a slight improvement in F1-score by 0.02. For both the LSTM and RF model, SHAP and MDI analysis showed human inhibition factors had a substantial influence on the model's predictions; however, they were secondary to historical wildfire frequency and environmental (vegetation indexes, soil moisture) variables, indicating that human inhibition factors act best as a complement to these driving factors.

Wildfire risk probabilities outputted by the model showed concentrated risk in central and southwestern California and during time step 2 (May, June, July, August), with both spatial and temporal patterns matching historical trends exhibited in real life. Through this study, future researchers should use human inhibition factors to enhance wildfire risk prediction models, minimizing casualties, economic loss and environmental downsides.

--Introduction--

Wildfires lead to loss of human life, property, biodiversity, air quality, and government financial liquidity (Thomas, 2017). Furthermore, they release harmful greenhouse gases which contribute to climate change, leading to a positive feedback loop with wildfires growing in frequency and severity in recent years (Xu, 2024). Just in California, the average rate of acres burned per year from 2009-2018 was 708 thousand compared to 337 thousand acres between 1979 and 1988, and the structural losses caused by wildfires in the year 2018 reached a record-breaking high of upwards of 4 billion dollars (Buechi, 2021).

Thus, wildfire risk prediction is important as it can help guide fire risk-management measures that governments can take to reduce the negative consequences of wildfires (Papakosta, 2017). Through risk prediction, fire managers can have a clearer understanding of how risks are spread in the areas they manage and guide fire risk-management measures that governments can take to reduce the negative consequences of wildfires such as fuel management planning, changes in response capacity, prioritization of high-risk areas in strategic incident response strategies, and increasing funding of selective ignition prevention programs (Scott et.al, 2013).

The prevalent factors that are included in current wildfire prediction models are largely landscape variables like elevation, fuel variables like vegetation indexes, meteorological features like temperature or evapotranspiration, and anthropogenic variables such as presence of powerlines or population density (Jain, 2020; Xu, 2024). Currently, there is a lack of inclusion of explicit human inhibition factors in currently-used models which may lead to inaccuracies regarding the impact of humans on lowering wildfire risk such as the presence of firefighters (Xu, 2024). Furthermore, since highly populated and lowly populated areas are at the lowest risk of wildfires with sparsely populated areas being the most likely areas for wildfires to take place, the presence of humans may have a mitigating effect on the presence of wildfires that is not presently included in current wildfire risk prediction models (Butsic, 2015).

Background Information:

Wildfires and methods used for risk prediction

Wildfires start from natural (e.g. lightning strikes) or human-caused ignitions (e.g. power lines) which are influenced by socioeconomic and biophysical spatial variability, and the intensity and likelihood of that fire ignition is shaped from fuel, topography, weather, and terrain variables (Scott et.al, 2023). Due to the complex nature of wildfires, with inaccurate or the absence of risk prediction models, wildfire management efforts are likely to be less impactful and often lead to wasteful use of resources (Scott et.al, 2023).

Historically, statistical and GIS models were the norm until the late 2000s which predicted fire risk by identifying correlations between mainly historical fire occurrences, environmental, meteorological variables and occasionally social variables (e.g., temperature, humidity, human proximity) using satellite remote sensing data or historical records, often using logistic regression to estimate fire probabilities in a given area based on historical fire data (Zacharakis, 2023). However, since these models relied on predetermined expert rules on how different factors affected wildfire risk and assumed that wildfire phenomena can be represented by linear quantifiable relationships, the statistical models were unable to generalize over different regions and were limited in the number of features they could consider. Moreover, the GIS models couldn't treat relationships between features as complex or non-linear which is closer to reality. (Zacharakis, 2023).

Currently, advancements in computational performance, sensor resolution, and storage capacity have dramatically increased the volume of data and input parameters available for analysis. While traditional statistical methods struggle to manage this complexity, machine learning techniques have been shown to effectively handle large amounts of data (Zacharakis, 2023).

Machine Learning (ML) Models

Machine learning is a branch of artificial intelligence (AI) that focuses on creating models that learn underlying patterns of data without needing extensive expert rules or precise environmental modeling. It uses statistical methods like regression, classification, clustering, or dimensionality reduction to find relationships in data, optimizing performance through criteria like accuracy or error minimization (Murphy, 2012).

A common type of ML models is supervised learning where models are trained on pre labeled datasets in which each data point can be mapped to a known classification through which the model learns how to map inputs to the correct outputs. After training, the model is tested on by labeling a new unlabeled dataset and its performance is evaluated by comparing its predictions to the known classifications in the testing data (Jain, 2020).

Deep Learning Models

Deep learning algorithms are a subset of ML algorithms that use multiple layers of neural networks to automatically extract and understand different levels of patterns. The initial layers learn simpler patterns and pass on their refined output to the layers after to progressively find more complex patterns and then feed this input to the next layers and so on to best model the raw data, without requiring manual feature engineering such as scaling, normalization, feature selection (removing unimportant features) and feature construction (creating new features out of other features such as body mass index from weight and height) which are often necessary for ML models to perform well on datasets (LeCun, 2015). In this study, RFs are compared with the performance of the DL model.

Long Short-Term Memory (LSTM) Model

Long Short-Term Memory (LSTM) networks are a DL model that maintains and learns from long-term dependencies within sequential data. Since the spatial influence of the wildfire tiles are being recorded through averages of neighboring tiles and collected in tabulated form, the problem transitions into a time-series analysis, making LSTMs a suitable approach for modeling these temporal relationships.They use a series of gates (input, forget, and output gates) to control the flow of information. The input gate decides what new information to store, the forget gate determines what information to discard, and the output gate selects what to pass to the next time step. Through this, LSTMs effectively maintain long-term dependencies and adapt to new patterns in sequential data (LeCun, 2015). Due to their ability to deal with time series well, they are useful for wildfire risk prediction as they can better analyze the temporal dependencies between numerous interconnected variables thus improving the prediction of fire risk based on current and past conditions (Xu, 2024).

Literature Review

Current ML and DL models fall short in their ability to accurately predict wildfire risk and are often skewed in favor of over-predicting wildfires.

For example, a wildfire occurrence prediction model covering Italy using a Random Forest framework developed by Cilli et al. (2022) exhibited a precision of 50.9%. Its inputs consisted mainly of the fire weather index (general index based on fuel moisture and fire behavior), vegetation data, terrain data, and neutral human variables and it outputted a binary value of whether a wildfire occurred or not. The low precision of the model indicates a high number of false positives (meaning that the model predicted presence of a wildfire where there was none) and, consequently, overestimated the presence of wildfires (Cilli et al., 2022). Furthermore, a meta-analysis study showed that when studies balanced their wildfire dataset with a nearly 50% split, the model predicted far more false positives as there were far more example of wildfires occurring in the training dataset than the testing dataset (Xu, 2024). Lastly, in a statistical study that used historical wildfire data from Oregon, the maximum probability of a wildfire occurring over every grid in Oregon was 0.0004 (Preisler, 2004). This probability is significantly lower than estimates from many deep learning models, which often predict wildfire risk probabilities in the range of tenths, suggesting that DL models are overestimating wildfire risk (Kondylatos et al., 2022).

Current ML and DL models in wildfire risk prediction lack explicit human inhibition factors that could contribute to false alarm errors.

Studies often include human inhibition factors implicitly, such as in Mann et al. (2016) which analyzed California's population density and its influence on wildfire risk, and Mansuy et al. (2019) which utilized the human footprint index (HFI) in U.S. protected areas. These studies used variables indirectly related to human inhibition factors, whereas this study will focus on explicit variables directly linked to wildfire inhibition, such as the presence of firefighting facilities and fuel treatment acres, thus providing a clearer picture of their impact. In rare cases where studies included explicit human inhibition factors, they were typically limited to a single metric, such as distance from firebases. For instance, Robinne et al. (2016) analyzed the effect of anthropogenic factors on wildfire risk in Alberta, Canada using statistical methods, but only considered one explicit human inhibition factor.

In that regression model, adding distance to fire fighting facilities as a factor led to improvements in the model's performance, adding a significant gain of 1.34% in deviance. Through this, if more explicit human inhibition factors are included such as acres impacted by fuel reduction projects (vegetation management to reduce wildfire impact) or funding allocated to facilities, wildfire risk prediction models can be improved (Robinne et al., 2016).

Furthermore, as wildfire risk prediction models trend towards ML or DL, the usage of human factors has largely been focused on human-induced ignition of wildfires, leaving out the mitigating influence of firefighting infrastructure (Xu, 2024).

Deep learning models outperform ML and statistical alternatives, and specifically, SHAP analysis can be used to quantify the impact of human inhibition factors on wildfire risk.

For example, in Xu et. al (2024), a technical meta-analysis that covers the current state of wildfire risk prediction states that for a majority of papers, DL methods often outperform ML methods, as they can handle large swathes of data more effectively, and can more effectively learn real-world conditions by forming non-linear relationships. Furthermore, in another meta-analysis, Ghali (2023) reviewed recent deep-learning approaches for detecting, mapping, and predicting wildfires, and evaluated the effectiveness of these use cases. From this, it was found that DL outperforms traditional ML methods in wildfire detection and mapping, such as the DL model proposed by Omar et.al in 2021 which had a root mean squared error (RMSE) of 0.021 and performed better compared to more traditional ML models such as decision trees with an RMSE of 0.274 (Omar et.al, 2021). Thus, for this study, DL was chosen as it has the highest accuracy of existing wildfire prediction models and due to its recency lacks explicit human inhibition factors.

Lastly, in Shmuel, et.al (2023), a study that developed a model to predict daily wildfire growth rates using machine learning models, SHAP was used to quantify feature importance like previous wildfire behavior and meteorological conditions. Through this, it shows that SHAP provides a clear methodology for understanding model decision-making and that SHAP can be similarly applied to evaluate the impact of human inhibition factors.

Research Goals

1. Develop a more accurate deep-learning model for wildfire risk prediction utilizing novel explicit human inhibition factors.

2. Utilize feature analysis (SHAP and MDI) to better visualize the impact of specific inhibitors on wildfire risk.

--Methodology--

Data Acquisition

Input variables were classified into fuel conditions, meteorology and climate conditions, socioeconomic factors, terrain variables, hydrological features, and wildfire historical records for wildfire risk prediction models. These were set to a 4 month resolution, except for terrain variables (elevation, slope) or facility data, which were time-invariant. From these classifications, journal articles focusing on human impact used variables for climate and meteorological conditions, such as average annual evapotranspiration (mm), terrain conditions such as slope (°) or elevation (m), the history of wildfires in that area, and most importantly, socioeconomic factors such as distance from human settlements and maximum housing density (Mann, 2016). In other studies, other meteorological variables were used, such as temperature, humidity, precipitation, wind speed, and vegetation data in the form of NDVI and EVI vegetation indexes (Malik et al., 2021). Furthermore, as human inhibition factors were not often analyzed in studies, there was a lack of data analysis on fuel reduction treatments, which were sourced from CAL-Fire's publicly available data (Xu, 2024). Additionally, each tile contained the number of firefighting facilities and the average number of firefighting stations in neighboring tiles. Further details about the input variables are shown in Figure 1.

Variable	Source	Description (Units)	Time Variant
NDVI	MODIS	Normalized Difference Vegetation Index	True
EVI	MODIS	Enhanced Vegetation Index	True
Net Evapotranspiration	MODIS	Evapotranspiration (mm/month)	True
Potential Evapotranspiration	MODIS	Potential Evapotranspiration (mm/month)	True
Weather data (ERA5)	ERA5	ERA5 Satellite Weather Data	True
Potential evaporation	ERA5	Potential evaporation (mm)	True
100m U component of wind	ERA5	100m U wind component (m/s)	True
Total precipitation	ERA5	Total precipitation (mm/month)	True
2m Temperature	ERA5	Temperature at 2m (°C)	True
Surface net solar radiation	ERA5	Net solar radiation (W/m^2)	True
100m V component of wind	ERA5	100m V wind component (m/s)	True
Surface runoff	ERA5	Surface runoff (mm/month)	True
Terrain slope (°)	SRTM	Terrain slope (°)	False
Terrain elevation (m)	SRTM	Terrain elevation (m)	False
# of wildfires	CAL-FIRE	Wildfire count	True
Facility Count	CAL-FIRE	# of active firefighting facilities	False
Acres of Fuel Treatment	CAL-FIRE	Acres annual	True
# of Fuel Treatments	CAL-FIRE	# of Fuel Treatments	True
Soil Moisture Surface	SMAP	Surface soil moisture (mm)	True
Soil Moisture Root	SMAP	Rootzone soil moisture (mm)	True
Soil Moisture Profile	SMAP	Profile soil moisture (mm)	True
Imperviousness	USGS NLCD	Impervious surface percentage	False

Figure 1: List of features, sources, description in units and time variance. Generated by student author

Data Processing

California was then split into 76 tiles with a 4 month temporal resolution, fuel conditions, meteorology, and climate conditions were evaluated. For example, for each tile, every 4 months, the median normalized vegetation index (NDVI) was calculated and stored. The training dataset included 2018-2021 vegetation data, while every other feature was from 2019-2022, varying by year and tile. Data that were time-invariant, such as elevation and the number of facilities, remained unchanged. The testing dataset was set one year later, including 2022 vegetation data, and everything else was set to 2023, excluding the number of wildfires. The prediction for the number of wildfires in 2023 per tile per year was then compared to the actual number of wildfires in 2023 per tile per year. Unless it was a variable that represented a count or used median (NDVI, EVI), features were aggregated using the mean of values inside the tile during the 4 month time step.

Model Construction

Since RNNs (recurrent neural networks) are built for sequential data, they are most useful for assessing and validating the risk prediction model using historical wildfire data (Xu, 2024). A variant of RNNs, LSTM (Long Short-Term Memory), was used, as it was better suited for long-term prediction tasks and could selectively forget information if it significantly reduced accuracy. Furthermore, it was shown to have the highest accuracy compared to traditional RNNs when assessing wildfire risk with historical wildfire and meteorological data in past literature (Liang, 2019). As a point of reference, random forest (RF) was trained on the input variables, with ground truth as the actual number of wildfires that occurred per tile per year in 2023, using the cost function of mean squared error and an Adam optimizer.

Validation of Model: Metrics and Predictions

The model will be validated based on historical wildfire data and comparing how the risk predictions vary from ground truth. The model will then be compared through the utilization of a receiving operating characteristic (ROC) curve, a precision-recall (PR) curve, macro F1-score, macro precision, macro recall, and accuracy (Liang, 2019).

Precision is the proportion of the model's predicted positive observations (true positives and false positives) that are correctly identified as positive (true positives), measuring the accuracy of the model's positive predictions (Foody, 2023). Recall is the ratio of correctly predicted positive observations (true positives) to the actual positive values (true positives and false negatives), indicating whether the model can correctly identify all objects of the target class, even with incorrect predictions (Foody, 2023). From precision and recall, F1-score is calculated as the harmonic mean of both values, combining them into a single metric through which a model can be assessed and compared to others in classification tasks (Foody, 2023). To ensure balanced accuracy across classes, these metrics were aggregated using a macro-average which computes the metrics independently for both classes (no wildfire and wildfire), and then averages them (Takahashi et.al, 2021). Accuracy is the ratio of correctly predicted observations (true positives and true negatives) to the total observations, measuring how often the classifier makes correct predictions (Foody, 2023).

AUROC (area under the receiver-operating characteristic curve) is the area under a graph that compares how many actual positive cases the model identifies (recall) and how many positive cases the model incorrectly identifies (false positive rate) across different decision thresholds (the value the probability needs to exceed for a sample to be classified as positive) (Kondylatos, 2022). AUPR (area under the precision-recall curve) is the area under a graph that represents the relationship between precision and recall (Keilwagen, 2014). For AUPR and AUROC, the closer it is to 1, the better the model's performance (Keilwagen, 2014; Kondylatos, 2022).

Furthermore, feature weighting and normalization will be done to quantify the impact of each human inhibition factor on wildfire risk which can be used to guide better policy and organization of resources. This was primarily done through the use of the python library SHAP for the LSTM, a commonly used program that can explain ML models and a variety of DL models and return the importance of features within the dataset (Lundberg, 2017). In addition to SHAP, random forest's MDI (mean decrease in impurity) (which computes the contribution of a feature to reducing uncertainty in a model's predictions) was used to further quantify the impact that human inhibition factors had on the model.

--Results--

The current state-of-the-art wildfire prediction models overpredict wildfires. In this study, we presume that the novel addition of human inhibition factors (average number of fire fighting facilities in surrounding tiles, number of firefighting facilities, acres treated to reduce wildfire risk, number of land management projects to reduce wildfire risk) will decrease the likelihood of false positive predicted results. To evaluate this, we compared 2 algorithms (LSTM and RF), both with and without the inclusion of these inhibition factors, and compared performance metrics. We then performed feature ranking to visualize which factors most informed the models' respective predictions.

Both Random Forest and LSTM models showed high predictive success

As both LSTM and Random Forest had macro F1-score averages in the mid 0.8 range (with the notable exception of LSTM without Human Inhibition Factors), they show clear success in the prediction of both the absence and presence of wildfires (Figure 2). Furthermore, in all cases, the accuracy of the models are above 80%, indicating that 80% of predictions are correct (Figure 2). As a point of comparison, since there were 152 instances of wildfire absence and 76 instances of wildfire presence in the test dataset, a theoretical model that only guessed 0 or 1, would have an accuracy of 0.67 or 0.33, respectively. Thus, the high values of these metrics for the LSTM model with Human Inhibition Factors and the Random Forest models indicate that these models are robust and effective predictors of wildfire occurrence.

Incorporating human inhibition factors improves the efficacy of the wildfire risk prediction LSTM model, and marginally improves the Random Forest (RF) model.

For the LSTM model, the inclusion of human inhibition factors improves all of its metrics (macro F1-score, macro recall, macro precision and accuracy). For macro precision, there is an increase from 0.82 to 0.86, indicating that the model is less likely to make incorrect classifications (Figure 2). For macro recall, the increase from 0.77 to 0.85 signifies that the model is more sensitive to wildfire occurrences or absences (Figure 2). This culminates in the increase of macro F1-score from 0.78 to 0.85 and in the increase of accuracy from 0.82 to 0.87, showcasing that human inhibition factors for the LSTM model increased its effectiveness (Figure 2). Through this, the LSTM model with human inhibition factors demonstrates greater overall correctness in its predictions, making it more reliable in distinguishing between the presence and absence of wildfires.

LSTM	Without Human Inhibition Factors	With Human Inhibition Factors
Precision	0.82	0.86
Recall	0.77	0.85
F1-score	0.78	0.85
Accuracy	0.82	0.87
Random Forest	Without Human Inhibition Factors	With Human Inhibition Factors
Precision	0.84	0.87
Recall	0.86	0.86
F1-score	0.84	0.86
Accuracy	0.86	0.88

Figure 2: Evaluative Metrics for Wildfire Risk Prediction Model. LSTM and RF model was used for a binary classification problem where presence of wildfire counted as a 1 and lack of wildfires as 0. Precision is correctly predicted positive observations divided by the total predicted positive observations, recall is correctly predicted positive observations divided by all observations in the actual class, f1-score is the harmonic mean between both precision and recall, and accuracy is the percentage of correct predictions to all predictions. Here, macro precision, macro recall, and macro F1-score averages were used, meaning that precision, recall, and F1-score were calculated separately for each class (wildfire presence and absence) and then averaged, giving equal weight to both classes. Macro averages were used to deal with the unbalanced nature of the dataset and to give equal weight to the correct prediction of absence and presence of wildfires. Generated by student author

The Random Forest model showed marginal improvement, mainly in macro precision as recall remained equal at 0.86. Similar to LSTM, macro precision increased from 0.84 to 0.87 (Figure 2). The overall F1-score also increased slightly from 0.84 to 0.86, and accuracy increased slightly from 0.86 to 0.88 further highlighting the model's enhanced reliability in classifying wildfire events accurately with the addition of human inhibition factors (Figure 2).

Inclusion of inhibition factors produced opposite shifts in sensitivity to actual wildfire events: RF became less sensitive while LSTM became more sensitive to true positives (correct prediction of presence of a wildfire).

With the inclusion of human inhibition factors, the LSTM model decreased in false negatives (prediction of no wildfire when there is a wildfire) from 30 to 17 (Figure 3). Since the test dataset had 228 instances, this decrease in false negatives represents a 5.7% improvement in identifying wildfires.

Through this, the inclusion of human inhibition factors for LSTM further improves the model, mainly by enhancing the model's sensitivity to actual wildfire events.

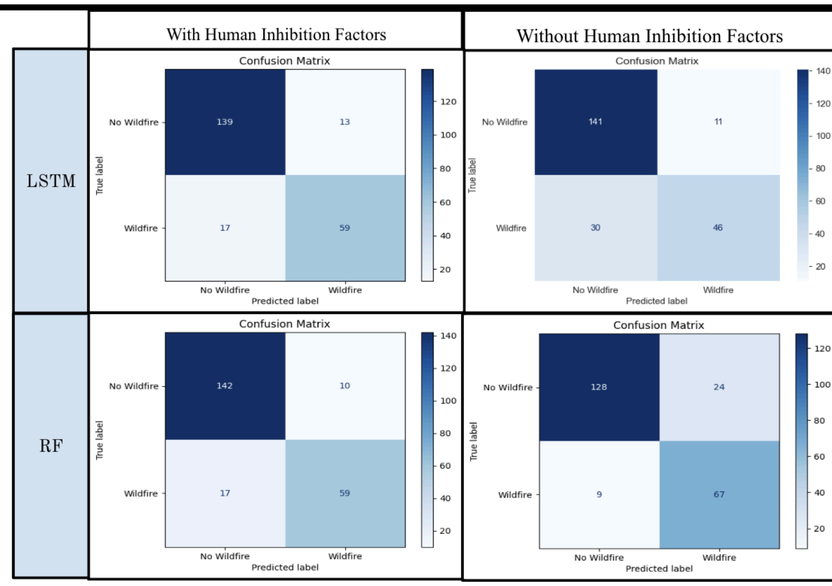

Figure 3: Confusion Matrices with and without Human Inhibition Factors for LSTM and Random Forest. Confusion matrices are 2x2 grid which represent the number of predictions that a model makes which are either true positives (model correctly predicted that a wildfire would occur), true negatives (correctly predicted that a wildfire would not occur), false negatives (model incorrectly predicted the absence of a wildfire), and false positives (model incorrectly predicted the presence of a wildfire). In the case of 2023 in 4-month intervals over 76 tiles, there were 152 instances where wildfires did not occur and 76 instances where wildfires did occur. (Generated by student author)

In contrast, human inhibition factors made the RF model more conservative and increased false negatives from 9 to 17, a 3.5% increase in misclassifications (Figure 3). This shift could stem from the inclusion of additional features altering the importance assigned to existing predictors. By relying more on the new human inhibition factors, the model may potentially diminish the influence of other critical variables, and thereby affect its sensitivity to actual wildfire events.

The inclusion of human inhibition factors reduced false positives in the RF model; however, LSTM saw little change in its false positives (predicting wildfire where there is none).

For the RF model, with human inhibition factors, false positives reduced from 24 to 10 (Figure 3). This 6% decrease in false positives further indicates that the inclusion of inhibition factors makes the RF model more conservative, underpredicting wildfire occurrences and thus being less likely to predict a wildfire occurring incorrectly.

For the LSTM, there is an incredibly slight increase (from 11 to 13) in false positives with the inclusion of human inhibition factors. This 0.9% increase is essentially negligible compared to the 228 instances in the testing dataset.

The AUROC and AUPR metrics demonstrate the high predictive success of the wildfire risk prediction models.

Further evaluation of the model through ROC curves was conducted in order to evaluate the discriminatory ability of the model in predicting the presence or absence of a wildfire with and without human factors. A higher AUROC (Area Under Curve of ROC) value means the model is better at separating the two classes (wildfire risk/ no wildfire risk), with a value of 1 being perfect and 0.5 being as good as random guessing (Bradley, 1997). The LSTM model with human inhibition factors achieved an AUROC of 0.90, and the RF model achieved 0.89. indicating high discriminatory ability between presence and lack of wildfire (Figure 4).

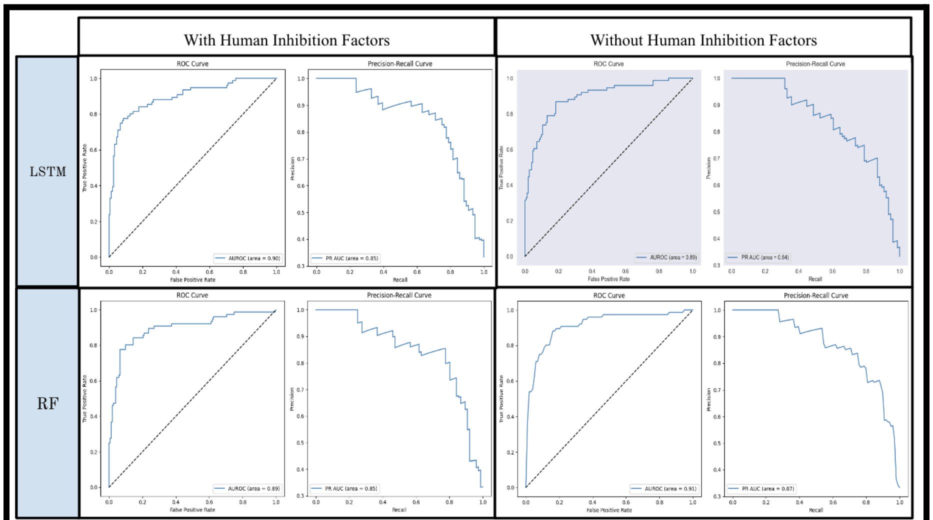

Figure 4: ROC and Precision-Recall Curves Comparing Model Performance with and without Human Inhibition Factors. ROC (Receiver Operating Characteristic) curves demonstrate the trade-off between true positive rate (recall) and false positive rate (incorrectly classified negatives over actual negatives) across different decision thresholds. In this case, the right panel shows the ROC curve for the model without human inhibition factors, with an AUC (Area Under Curve) of 0.88, while the left panel shows the ROC curve for the model incorporating human inhibition factors, with an AUC of 0.90. A higher AUC indicates better model performance in distinguishing between classes. Similarly the PR AUC (Area under Curve for Precision Recall) is 0.85 with the human inhibition factors and 0.84 without the human inhibition factors and a higher AUC indicates better model performance. (Generated by student author)

However, AUROC can be overly optimistic when dealing with imbalanced datasets as the dilution of the majority class can exaggerate the performance of the model in identifying the minority class. Since the percentage of positive values in our dataset was around 33%, indicating a slight imbalance, AUPR was used to provide a more accurate metric for the performance of the model specifically for identifying the presence of a wildfire occurring at varying thresholds.

Since the AUPR of the LSTM model with human inhibition factors and the RF with human inhibition factors are both a high score of 0.85 shows that both models are effective and can accurately predict the chance of a wildfire occurring (Figure 4). In comparison, a model that randomly guessed would have an AUPR of around 0.33 (the percentage of positive values in the testing dataset).

With human inhibition factors, LSTM had a marginal increase and RF had a slight decrease in performance, specifically regarding the AUROC and AUPR metrics.

The LSTM model with human inhibition factors has an AUROC of 0.90 and AUPR of 0.85, in comparison to its AUROC of 0.89 and AUPR of 0.84 without them (Figure 4). Through this slight increase, the inclusion of human inhibition factors led to slight improvement of the LSTM model, particularly in its discriminatory ability and identifying presence of wildfires.

In contrast, the RF model's AUROC and AUPR decrease from 0.91 to 0.89 and 0.87 to 0.85, respectively, with the addition of human inhibition factors. This indicates that human inhibition factors led to a slight decrease in performance of the RF model, contrary to the LSTM's performance.

This decrease in RF's performance may be due to its static nature, as it does not account for temporal dependencies like LSTM. Since treatment acres and treatment count vary over time, RF lacks the ability to leverage their temporal variance, treating them redundant features that would dilute the model's predictive power.

Spatial and Temporal Case Study: California, 2023

In order to evaluate sequential trends, LSTM is the algorithm typically structured in best alignment with this goal based on its ability to capture and model long-term dependencies in time-series data. Due to it achieving a reasonably high threshold of performance metrics (macro F1-score of 0.85, AUPR score of 0.85), it was further levergated to evaluate spatial and temporal trends surrounding wildfire risk. The spatial and temporal distribution of wildfire probability for 2023 is illustrated by comparing the LSTM (with human inhibition factors) model's confidence that a given input instance falls into the positive class (a wildfire occuring) (Figure 6).

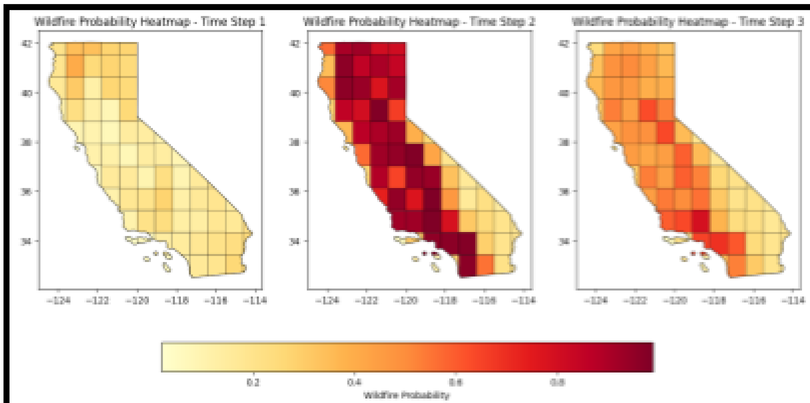

Figure 6: Prediction of Wildfire Occurring Probability Map. LSTM model with Human Inhibition Factors was used for a binary classification problem where presence of wildfire counted as a 1 and lack of wildfires as 0. The prediction of wildfires was for the year 2023 over 76 tiles in California and 3 time steps (4 months). (Generated by student author)

Wildfire probability peaks in the mid-year tri-section, with increasing risk from the first to second periods, then decreasing by the third, likely reflecting seasonal drying patterns.

In time step 1, during mainly winter months, the majority of tiles are light yellow, which suggests wildfire probabilities close to the lower end of the scale, approximately between 0.2 and 0.3.

From time step 1 to time step 2, there is a dramatic increase in the probability of a wildfire occurring with an average increase of around 0.2-0.3 and some high-risk regions experiencing an increase up to 0.4. This leads to final values in time step 2 in high-risk regions as around or above 0.8.

From time step 2 to time step 3, there is a slight decrease with an average reduction of about 0.1-0.2, and final values around 0.6-0.7 in high risk areas and the average probability to stabilize around 0.4.

The disparity between time steps is likely due to environmental factors (e.g., early winter rains or cooler temperatures) starting to reduce fire risk, especially in regions where risk was highest previously. Through this, the model prediction aligns with California's seasonal cycle, where wildfire probability is virtually nonexistent during the beginning of the year, peaks in the mid-year (late summer/fall) and slightly decreases towards the end of the year.

Central and southwestern California exhibit higher wildfire probabilities, especially in mid-latitudes (34-38), while northwestern and southeastern regions consistently show lower risk comparatively.

In central and southwestern California (longitudes -122 to -118, latitudes 35 to 40), the wildfire probabilities predicted are consistently higher in later time steps with a range of around 0.7 - 0.9 during time step 2 and 0.5-0.7 during time step 3, indicating that these tiles were most prone to wildfire risk.

As the southeastern region (latitudes 33-35, longitudes -114 to -118) of California's climate consists mostly of desert (a region with sparse vegetation and low precipitation), the wildfire risk prediction model accurately predicts that it has a very low probability of a wildfire occurring. This pattern is most apparent in time step 3 where the average probability of the southeastern region is within 0.2 to 0.3, indicating a low chance of a wildfire occurring.

In northwestern California (latitudes 38–42, longitudes -122 to -124), across all time steps, the average probability of a wildfire occurring is around or under 0.5. Wetter and colder climatic conditions most likely aid in the reduction of wildfire risk in this region.

False negatives (over prediction of wildfire risk, shown in blue) are concentrated in the southeastern corner of time step 2

There is a high concentration of false negatives (predicted no wildfire when there actually was one) in the southeastern corner during time step 2 (Figure 7).

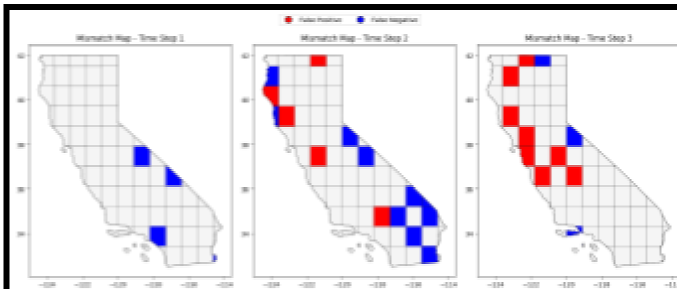

Figure 7: Mismatch Map that highlights errors made by LSTM Model. Displayed in the figure above is the false negatives and false positive errors that the model made spatially and temporally. This was created to provide further insight into what errors the model made, and the weaknesses of it in predicting wildfire risk. (Generated by student author)

This pattern indicates that the model may be overlooking some factors that are relevant to wildfire risk in southeastern California. For example, if past wildfires are less frequent in this area, the model may have learned fewer examples of wildfire-prone conditions specific to the region. Thus, specific factors aren't given more weight and the model doesn't understand the full impact of some variables on wildfire risk. As a result, it may underpredict fire risk here, especially since it is trained primarily on data from regions with different conditions.

Another possible reason is since southeastern California has sparse vegetation, it can lead to the model underpredicting wildfire risk if it gives too much weight to vegetation density and not other environmental factors.

Lastly, since the dataset effectively made the tiles uniform, localized anomalies were much harder to detect. Essentially if a tile had a mix of climates with varying wildfire risk, the model would have struggled to predict wildfire occurrences in those regions as the metrics considered would average the differences in climate.

False positives (over prediction of wildfire risk, shown in red) are mostly concentrated in time step 3 in the western part of California.

There is a high concentration of false positives (predicted wildfire where there actually was none) in the western region during time step 3 (Figure 7). Since western California has dense vegetation, it can lead to the model overpredicting wildfire risk if it gives too much weight to vegetation density. Furthermore, if the training data had overrepresentation of the frequency of wildfires in that region, it may develop a bias towards predicting a wildfire occurrence in that region, leading to false positives.

Human inhibition factors had a moderate impact on the prediction of the LSTM model.

The objective of SHAP analysis is to quantify the impact that an input variable has on the model's output predictions. Particularly treatment count (number of fuel reduction or land treatment activities done) and facility count (number of active fire stations, communication stations, lookouts, air attack bases, and Helitack bases per tile) were shown to have moderate impact on the models' predictions with mean Shapley values of around 0.022 for treatment count and 0.021 for facility count (Figure 8). Furthermore, they were among the top ranking features used in the model.

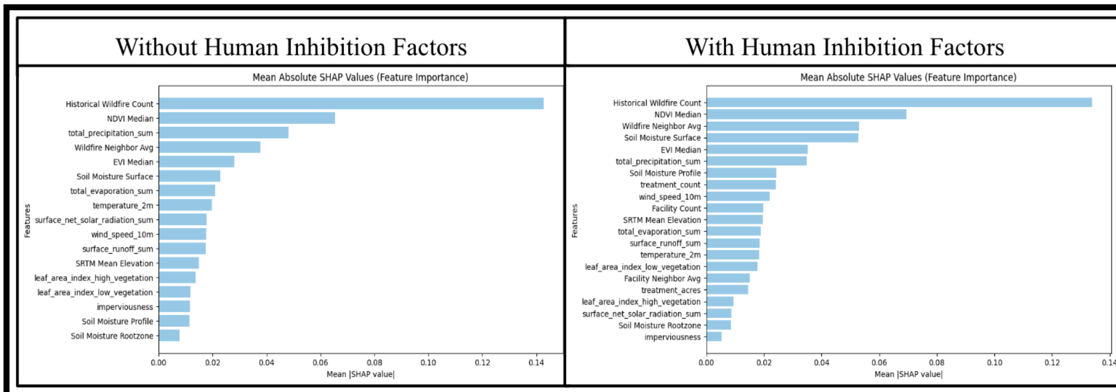

Figure 8: Feature importance of LSTM using Mean SHapley Values with and without Human Inhibition Factors. SHAP analysis was done on the LSTM model with and without human inhibition factors. The mean SHAP value is displayed above, indicating the average contribution of each feature to the model's predictions, derived from a sample of instances from the dataset. This was done using the KernelExplainer method. (Generated by student author)

The other human inhibition factors included facility neighbor avg (average number of active fire stations, communication stations, lookouts, air attack bases, and Helitack bases in neighboring tiles) and treatment acres (acres affected by land treatment activities to reduce wildfire risk) had slight impact. With mean Shapley values of around 0.018 and ranking in the bottom half of features, they provided a marginal supplementary benefit to the model's predictions.

The prediction of wildfires remains predominantly driven by historical wildfire data and environmental conditions.

Historical wildfire count (total wildfires that occurred in that tile from 2010 to 2016 during the specific quarter of the year) with a Shapley mean value of nearly 0.138 remains the main driving factor behind the model's predictions even with the addition of human inhibition factors (Figure 8). NDVI median (vegetation index) with a Shapley mean value of around 0.071, and other environmental features such as soil moisture surface (moistness of top layer of soil) and precipitation remain among the top influencing factors in the models' predictions (Figure 8) with Shapley mean values around 0.057 and 0.049, respectively.

Through this, human inhibition factors play a complementary role in relation to the more essential environmental features, in the context of the LSTM model.

From the SHAP analysis, "Facility count" and "treatment count" had positive correlations with wildfire risk, "facility neighbor average" had a negative correlation, and "treatment acres" was neutral.

From the SHAP analysis, a beeswarm plot was made to represent individual instances and provide greater insight into the general directionality of the SHAP importance.

Treatment count and facility count both show clustering with low values around -0.1 to -0.03 SHAP value and high values having a high positive impact on the models' predictions with values around 0.2 SHAP (Figure 9). Through this, lower values have negative to neutral impact on the model's predictions and higher values have very high positive impact on the model's predictions, indicating a positive correlation.

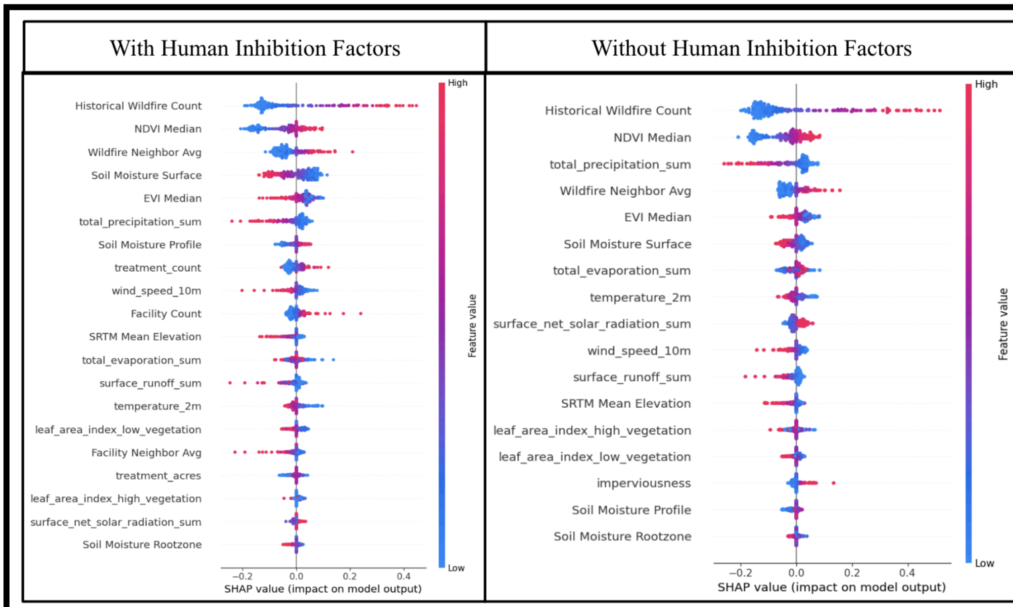

Figure 9: Average SHAPely rankings (estimated feature weights) of the model's inputs. Feature weights were evaluated using the SHAP library to quantify the impact of human inhibition and other input variables of the model. Shapley values quantify each variable's contribution to the model's predictions. Each dot represents a positive test example and is colored based on the scaled value of the corresponding variable. The input features are ranked based on absolute mean SHAP value. (Generated by student author)

Facility neighbor average shows that for high values, the SHAP values tend to be highly negative with around -0.05 to -0.2 SHAP value (Figure 9). For lower values, the instances concentrate around 0.03 and are mostly neutral or slightly positive (Figure 9). From this, facility neighbor average has a negative correlation with wildfire risk.

High values of treatment acres are heavily concentrated at 0 with low values having both slightly negative and slightly positive SHAP values. This indicates that this feature is largely neutral and has minimal influence on the model's predictions. Thus, the presence or absence of this feature should not greatly affect wildfire risk predictions for the LSTM model.

Human inhibition factors were among the most important features for RF.

Human inhibition features were crucial for RF's predictions. When all features were considered, facility count ranked highest, with treatment acres and count among the top 10 features in the model (Figure 10).

Figure 10: Top 10 features with largest feature importance for Random Forest Algorithm with and without Human Inhibition Factors. Upon identifying features that did not contribute strongly to the wildfire risk prediction model, the top 10 features were selected. The model was rerun, with the above shifts in feature importance. Calculated using Mean Decrease in Impurity. (Generated by student author)

When considering only the top 10 features, treatment acres, treatment count, and historical wildfire count's increased in importance to the top 3 features while facility count's influence decreased dramatically, suggesting that facility count plays a more supportive role when compared to fuel reduction, environmental and historical factors in RF. Thus, human inhibition factors when included impacted the model's predictions heavily in both cases when all or only the top 10 features were considered.

--Discussion--

The objectives of this study were to develop a more accurate wildfire risk prediction model by taking into account human inhibition factors (e.g. # of firefighter facilities, fuel reduction projects, and to quantify their specific impact on wildfire damage and risk.

Both the LSTM and RF models with novel inclusion of human inhibition factors demonstrated accurate predictions for wildfire occurrence probabilities

The LSTM and RF models, incorporating human inhibition factors, demonstrated strong predictive ability, as shown by macro F1-scores of 0.85 and 0.86, accuracies of 87% and 88%, and AUROCs of 0.90 and 0.89 (Figure 2, 4). Both had AUPRs of 0.85 which validate the models' sensitivity to actual wildfire occurrences (Figure 4). From these metrics, the LSTM model and the RF model can accurately predict wildfire occurrence over a 4 month span and 76 regions of California.

The LSTM model here performed slightly below the LSTM model in other literature in macro F1-score and accuracy, but outperformed them in AUROC. For example, an LSTM model used to predict wildfire danger as a binary classification problem in Greece when tested on 2020 and 2021 had macro f1-scores of 0.858 and 0.910, accuracies of 87.8% and 92.1%, and AUROCs of 0.868 to 0.886, respectively (Kondylatos et.al, 2022). Thus, while the LSTM model with human inhibition factors is slightly less accurate and had a lower F1-score than current studies, it achieved a higher AUROC.

The RF model, incorporating human inhibition factors, demonstrates superior performance across all metrics when compared to previous studies. For example, a study that built a random forest classifier for Italy, inputted the fire weather index, vegetation, terrain, and neutral human variables, and outputted a binary wildfire occurrence prediction had a F1-score of 0.651, an accuracy of 72.9% and an AUROC of 0.841 (Cilli et al., 2022). Another study that had built a wildfire danger prediction model for Greece as a binary classification problem had for the years 2020 and 2021, macro F1-scores of 0.792 and 0.842, accuracies of 0.83 and 0.867, and AUROCs of 0.824 and 0.870 (Kondylatos et.al, 2022). Thus, the RF model with human inhibition factors surpasses other models in literature in predictive ability.

Surprisingly, RF outperformed LSTM in almost all metrics. As this study used time-series data, the LSTM model was expected to perform better based on its effectiveness with sequential data. Most likely, this discrepancy is caused by insufficient training data, which limited the LSTM's ability to learn temporal dependencies.

Human inhibition factors led to the improvement of the LSTM model

Human inhibition factors showed a large improvement of the performance of the LSTM model, as reflected in the improved macro F1-score from 0.78 to 0.85, improved macro recall from 0.77 to 0.85, improved macro precision from 0.82 to 0.86 and improved accuracy from 0.82 to 0.87 (Figure 2). Furthermore, the inclusion of explicit human inhibition factors led to a marginal increase in AUROC from 0.89 to 0.90 and AUPR from 0.84 to 0.85 for the LSTM model, demonstrating how human inhibition factors such as facility count and acres of fuel treatment led to an increase in the discriminatory ability of the model (Figure 4).

Interestingly, the inclusion of human inhibition factors for LSTM primarily improved performance due to the reduction of false negatives (Figure 3). This indicates that the model showed an increased sensitivity to wildfires, allowing it to better capture instances where wildfires would occur. By reducing false negatives, the model becomes more reliable for decision-makers who rely on these predictions for real-time resource deployment and risk mitigation strategies.

Therefore, future LSTM wildfire risk prediction models should explicitly include human inhibition factors to better account for anthropogenic impact on wildfire risk to better inform their predictions.

Random forest showed mixed impact from the addition of human inhibition factors

Random forest marginally improves in macro F1-score, macro precision, macro recall and accuracy with the addition of human inhibition factors.

Surprisingly, their inclusion led to a 3.5 % decrease in false positives and a 6% increase in false negatives. Through this, the impact of human mitigation factors on the RF model runs contrary to that of the LSTM model. While the LSTM model became more sensitive to wildfire occurrences, RF became more conservative and less likely to predict a wildfire occurring.

Oddly, when the metrics AUROC and AUPR are evaluated for RF, there is a slight decrease with the inclusion of human inhibition factors.

This discrepancy can be attributed to the thresholding process which finds the best confidence threshold possible for the maximum F1-score. Essentially, the F1-score metric shows the peak performance of the model at its best possible threshold (Tharwat, 2018). The inclusion of human inhibition factors improved the performance at that optimal point and since other metrics (precision, recall, accuracy) are dependent on that threshold, they improved as well.

On the other hand, AUROC and AUPR are threshold independent and quantify performance of a model at all thresholds (Tharwat, 2018). Thus, human inhibition factors caused the model to improve at the optimal point but may have caused the model's classification ability to drop at other thresholds.

Human inhibition factors had moderate influence on the output of the wildfire risk prediction models, and environmental and historical data remain primary drivers.

For LSTM, treatment count (number of fuel reduction or land treatment activities done) and facility count (number of active fire stations, communication stations, lookouts, air attack bases, and Helitack bases per tile) were shown to have moderate impact on the models' predictions and were ranked relatively highly among the top features by mean Shapley value (Figure 8).

Interestingly, human inhibition factors had varying correlations with wildfire risk. Facility count and treatment count had positive correlations, facility neighbor average had a negative correlation and treatment acres had a neutral correlation (Figure 9). The positive correlation between facility count and treatment count most likely occurred through an unaccounted-for hidden covariate in the model. The negative correlation with wildfire neighbor average indicates that if surrounding tiles have a high number of fire fighting facilities, it creates a buffer effect where the surrounding tile is less susceptible to wildfire risk, in the context of the LSTM model. Lastly, treatment acres' neutral correlation indicates that it has marginal benefit to the improvement of the model and isn't significantly related to wildfire occurrence.

For RF, when all features were included, facility count was the most important feature (importance score of 0.05) (Figure 10). Treatment acres and treatment count were among the top 10 features as well, indicating high influence over the model's predictions (importance score of 0.01 each) (Figure 10).

Then, after the model was limited to the top 10 features, facility count dropped to the bottom of feature importance relatively (importance score of 0.06) and both treatment acres and treatment count skyrocketed to the top 3 most influential features (importance scores around 0.11) (Figure 10).

Environmental factors remain key drivers of wildfire occurrence in both LSTM and RF, with and without human inhibition factors.

With and without human inhibition factors, historical data such as past wildfire count (total number of wildfires from 2010 to 2016), and environmental features such as NDVI (vegetation index), soil moisture surface (moisture of top layer of soil), and precipitation remain among the top influencing factors in the LSTM model.

Similarly in RF, historical wildfire count consistently ranked as the most important feature, both with and without the inclusion of human inhibition features, when considering only the top 10 features. Additionally, soil moisture surface, NDVI median and EVI median remained among the top 10 features when all features were considered and displayed a moderate impact on the predictions of the model when the model was limited to the top 10 features.

Thus, in both LSTM and RF, environmental and historical data remained essential features for accurate predictions. As such, human inhibition factors act best as a complement to environmental features, aiding the more consequential features.

Other studies that previously covered neutral anthropogenic factors showcased similar trends. For example, a study that looked at Mediterranean Europe and used random forest to predict fire occurrence found that precipitation, soil moisture and relative humidity were among the most important drivers for the model's predictions (Oliveira et.al, 2012). The model also showed that socioeconomic features (road density, population density) were significantly important to the model's predictions, albeit to a lesser degree (Oliveira et.al, 2012). Another study that covered southwest China and used XGBoost to predict wildfire probability found that meteorological factors ranked more strongly than socioeconomic factors (distance to roads and residential areas) (Quan et.al, 2023).

Wildfire predictions followed seasonal time patterns with the highest likelihood of wildfires occurring in the 2nd time step.

Peak seasons for California wildfires, in recent years, are concentrated in July and August (Li & Banerjee, 2021). This aligns with the concentration in time step 2 (May, June, July, August) that was outputted by the LSTM model with human inhibition factors, indicating that the model accurately perceived this temporal trend from the training data (Figure 7).

In California, the central and southwestern regions showed far higher probability of risk than the southeast and northwest.

Due to environmental conditions and human factors, the central and southwestern regions of California are at significantly higher risk of wildfire occurrence (Li & Banerjee, 2021). Comparatively, less flammable vegetation in the southeast and lower temperatures in the northwest leads to less frequent wildfires (Li & Banerjee, 2021). These trends are reflected in the probability map generated by the LSTM model incorporating human inhibition factors where the central and southwestern regions had demonstrably higher probabilities of wildfire occurrence than the southeast and northwest (Figure 7). Furthermore, this alignment indicates that the model accurately perceived the spatial trends within the training data.

False positives were largely concentrated in the western region of California during time step 3 and false negatives were most apparent in the southeastern region during time step 2.

The model's weaknesses were mostly concentrated in specific regions and time steps. False positives (prediction of a wildfire when none occurred) were largely on the west coast during time step 3, likely due to overrepresentation of wildfire occurrence in the training set in that region and time step (Figure 8). On the other hand, false negatives (predicting no wildfire where one occurred) were clustered in the southeastern region during time step 2 (Figure 8).

This phenomenon could be attributed to several factors: the underrepresentation of wildfires in the training set for that region, an overreliance on vegetation features (since the southeastern region of California has sparse vegetation, the model may underestimate wildfire risk), or the removal the meteorological and climatic variation within a tile, obscuring possible local hotspots.

Limitations:

Due to limited publicly-available data over a 4-month time-scale in the state of California, the effect of other human inhibition variables such as location of fire hydrants or firebreaks could not be properly quantified in this study. Furthermore, this study is using the Cal-Fire dataset which has its own limitations as it does not include the locations of federally owned facilities or federal fuel reduction projects. Lastly, since the data for fuel treatments was limited to 2019-2023, the temporal resolution of the entire dataset for both the LSTM and RF models had to be limited within that time span.

Furthermore, while the KernelExplainer method used for SHAP assumes independence, which could introduce some inaccuracies in the feature weight estimations, the use of the entire training dataset to calculate SHAP values ensures that the estimation is as accurate as possible given the data and model used.

Future Directions:

Given the significant impact of human inhibition factors on the model's output and its inclusion led to improvement in prediction at a 4 month scale in California, future research should explore their effects on wildfire risk in other wildfire-prone regions. Given that regions adopt drastically different approaches to wildfire risk reduction based on their own policies, and the effects of those efforts will vary based on their climate conditions, human inhibition will have a variable effect based on the region. For example, a model examining daily wildfire trends in Canada or monthly wildfire trends in Australia may have differing results in the impact of human inhibition factors and could offer insight into how different approaches affect different areas.

Beyond wildfire risk, the impact of human inhibition factors on burn area prediction could be implemented by including an additional target feature and using RF and LSTM for regression, instead of classification. Incorporating anthropogenic inhibition factors—such as distance from fire-fighting facilities—have led to marginal improvements for past regression models for predicting wildfire burn area (Robinne et al., 2016). By extending this study, additional human inhibition features can be analyzed such as fuel treatment data and facility frequency for the purposes of burn area prediction.

Lastly, since the inhibition impact of humans on wildfire risk is significant, more detailed spatial data should be collected and analyzed on fire hydrants, firebreaks, or specific community outreach and fire education programs to understand their relation to wildfire risk and what can be done to improve their impact.

Conclusion:

In this study, human inhibition factors were shown to improve both the LSTM and RF model, albeit in opposite directions. For the LSTM, the model became more sensitive with human inhibition factors reducing false negatives by 5.7% (prediction of no wildfire where there was one) whereas for RF, they increased false negatives by 3.5% and reduced false positives by 6% (predicting a wildfire where there was none). Through SHAP and MDI analysis, human inhibition factors were shown to have a moderate influence on both models' predictions, acting mainly as complements for more consequential environmental drivers. Therefore, since both LSTM and RF with human inhibition factors both displayed high performance metrics (F1-score, AUROC, AUPR, etc.) and the LSTM model's predictions followed both historical spatial and temporal trends closely, this study developed two accurate wildfire risk prediction models using novel human inhibition factors and effectively demonstrated that human inhibition factors led to improvements in model performance.

--References---

Abdollahi, A., & Pradhan, B. (2023). Explainable artificial intelligence (XAI) for interpreting the contributing factors feed into the wildfire susceptibility prediction model. The Science of the Total Environment, 879, 163004. https://doi.org/10.1016/j.scitotenv.2023.163004

Bach, P., Kurz, M. S., Chernozhukov, V., Spindler, M., & Klaassen, S. (2024). DoubleML: An object-oriented implementation of Double Machine Learning in R. Journal of Statistical Software, 108(3), 1–56. https://doi.org/10.18637/jss.v108.i03

Bakke, S. J., Wanders, N., Van Der Wiel, K., & Tallaksen, L. M. (2023). A data-driven model for Fennoscandian wildfire danger. Natural Hazards and Earth System Sciences, 23(1), 65–89. https://doi.org/10.5194/nhess-23-65-2023

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7), 1145–1159. https://doi.org/10.1016/s0031-3203(96)00142-2

Breiman, L., Friedman, J., Stone, C.J., and Olshen, R.A. 1984. Classification and regression trees. Chapman & Hall, New York.

Breiman, L. 2001. Statistical modeling: the two cultures. Stat. Sci. 16(3): 199–215. doi:10.1214/ss/1009213726

Buechi Hanna, Weber Paige, Heard Sarah, Cameron Dick, Plantinga Andrew J. (2021) Long-term trends in wildfire damages in California. International Journal of Wildland Fire 30, 757-762.

Butsic, Van, Kelly, M., & Moritz, M. (2015). Land use and wildfire: A review of local interactions and teleconnections. *Land*, *4*(1), 140-156. https://doi.org/10.3390/land4010140

Cilli, R., Elia, M., D'Este, M., Giannico, V., Amoroso, N., Lombardi, A., Pantaleo, E., Monaco, A., Sanesi, G., Tangaro, S., Bellotti, R., & Lafortezza, R. (2022). Explainable artificial intelligence (XAI) detects wildfire occurrence in the Mediterranean countries of Southern Europe. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-20347-9

Davis, K. T., Peeler, J., Fargione, J., Haugo, R. D., Metlen, K. L., Robles, M. D., & Woolley, T. (2024). Tamm review: A meta-analysis of thinning, prescribed fire, and wildfire effects on subsequent wildfire severity in conifer dominated forests of the western US. *Forest Ecology and Management*, *561*, 121885. https://doi.org/10.1016/j.foreco.2024.121885

Foody G. M. (2023). Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient. PloS one, 18(10), e0291908. https://doi.org/10.1371/journal.pone.0291908

Ghali, R., & Akhloufi, M. A. (2023). Deep learning approaches for wildland fires using satellite remote sensing data: Detection, mapping, and prediction. *Fire*, *6*(5), 192. https://doi.org/10.3390/fire6050192

Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., and Scholkopf, B. 1998. Support vector machines. IEEE Intell. Syst. Appl. 13(4): 18–28. doi:10.1109/5254.708428

Jain, P., Coogan, S. C., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. *Environmental Reviews*, *28*(4), 478-505. https://doi.org/10.1139/er-2020-0019

Keilwagen, J., Grosse, I., & Grau, J. (2014). Area under Precision-Recall Curves for Weighted and Unweighted Data. *PLoS ONE*, *9*(3), e92209. https://doi.org/10.1371/journal.pone.0092209

Kondylatos, S., Prapas, I., Ronco, M., Papoutsis, I., Camps‐Valls, G., Piles, M., Fernández‐Torres, M., & Carvalhais, N. (2022). Wildfire danger prediction and understanding with deep learning. Geophysical Research Letters, 49(17). https://doi.org/10.1029/2022gl099368

Li, S., & Banerjee, T. (2021). Spatial and temporal pattern of wildfires in California from 2000 to 2019. Scientific Reports, 11(1). https://doi.org/10.1038/s41598-021-88131-9

Liang, H., Zhang, M., & Wang, H. (2019). A neural network model for wildfire scale prediction using meteorological factors. *IEEE Access*, *7*, 176746–176755. https://doi.org/10.1109/access.2019.2957837

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017)*, 4765–4774. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316. https://doi.org/10.1097/jto.0b013e3181ec173d

Mann, M. L., Batllori, E., Moritz, M. A., Waller, E. K., Berck, P., Flint, A. L., Flint, L. E., & Dolfi, E. (2016). Incorporating anthropogenic influences into fire probability models: Effects of human activity and climate change on fire activity in california. *PLOS ONE*, *11*(4), e0153589. https://doi.org/10.1371/journal.pone.0153589

Mansuy, N., Miller, C., Parisien, M., Parks, S. A., Batllori, E., & Moritz, M. A. (2019). Contrasting human influences and macro-environmental factors on fire activity inside and outside protected areas of North America. *Environmental Research Letters*, *14*(6), 064007. https://doi.org/10.1088/1748-9326/ab1bc5

Murphy, K. 2012. Machine learning: a probabilistic perspective. MIT Press. Available from http://www.amazon.com/Machine-Learning-Probabilistic-PerspectiveComputation/dp/0262018020.

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., & Pereira, J. M. (2012). Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. Forest Ecology and Management, 275, 117–129. https://doi.org/10.1016/j.foreco.2012.03.003

Omar, N., Al-Zebari, A., & Sengur, A. (2021). Deep Learning Approach to Predict Forest Fires Using Meteorological Measurements. *IEEE*. https://doi.org/10.1109/iisec54230.2021.9672446

Papakosta, P., Xanthopoulos, G., & Straub, D. (2017). Probabilistic prediction of wildfire economic losses to housing in cyprus using bayesian network analysis. *International Journal of Wildland Fire*, *26*(1), 10. https://doi.org/10.1071/wf15113

Prapas, I., Ahuja, A., Kondylatos, S., Karasante, I., Panagiotou, E., Alonso, L., Davalas, C., Michail, D., Carvalhais, N., & Papoutsis, I. (2022). Deep learning for global wildfire forecasting. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2211.00534

Preisler, H. K., Brillinger, D. R., Burgan, R. E., & Benoit, J. W. (2004). Probability based models for estimation of wildfire risk. International Journal of Wildland Fire, 13(2), 133. https://doi.org/10.1071/wf02061

Quan, X., Wang, W., Xie, Q., He, B., De Dios, V. R., Yebra, M., Jiao, M., & Chen, R. (2023). Improving wildfire occurrence modelling by integrating time-series features of weather and fuel moisture content. *Environmental Modelling & Software*, *170*, 105840. https://doi.org/10.1016/j.envsoft.2023.105840

Robinne, F.-I., Parisien, M.-A., & Flannigan, M. (2016). Anthropogenic influence on wildfire activity in Alberta, Canada. International Journal of Wildland Fire, 25(11), 1131. https://doi.org/10.1071/wf16058

Sarker, I.H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN COMPUT. SCI. 2, 420 (2021). https://doi.org/10.1007/s42979-021-00815-1

Scott, Joe H.; Thompson, Matthew P.; Calkin, David E. 2013. A wildfire risk assessment framework for land and resource management. Gen. Tech. Rep. RMRS-GTR-315. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 83 p.

Shadrin, D., Illarionova, S., Gubanov, F., Evteeva, K., Mironenko, M., Levchunets, I., Belousov, R., & Burnaev, E. (2024). Wildfire spreading prediction using multimodal data and deep neural network approach. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-52821-x

Shmuel, A., & Heifetz, E. (2023). A machine-learning approach to predicting daily wildfire expansion rate. *Fire*, *6*(8), 319. https://doi.org/10.3390/fire6080319

Takahashi, K., Yamamoto, K., Kuchiba, A., & Koyama, T. (2021). Confidence interval for micro-averaged F1 and macro-averaged F1 scores. Applied Intelligence, 52(5), 4961–4972. https://doi.org/10.1007/s10489-021-02635-5

Thomas, D., Butry, D., Gilbert, S., Webb, D., & Fung, J. (2017, November). *The costs and losses of wildfires: A literature survey*. National Institute of Standards and Technology. https://doi.org/10.6028/nist.sp.1215

Tharwat, A. (2018). Classification assessment methods. Applied Computing and Informatics, 17(1), 168–192. https://doi.org/10.1016/j.aci.2018.08.003

Xi, D. D., Taylor, S. W., Woolford, D. G., & Dean, C. (2019). Statistical models of key components of wildfire risk. *Annual Review of Statistics and Its Application*, *6*(1), 197-222. https://doi.org/10.1146/annurev-statistics-031017-100450

Xu, Z., Li, J., & Xu, L. (2024). Wildfire risk prediction: A review. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2405.01607

Yann Lecun, Yoshua Bengio, Geoffrey Hinton. Deep learning. Nature, 2015, 521 (7553), pp.436-444. ff10.1038/nature14539ff. Ffhal-04206682f

Zacharakis, I., & Tsihrintzis, V. A. (2023). Integrated wildfire danger models and factors: A review. *The Science of the Total Environment*, *899*, 165704. https://doi.org/10.1016/j.scitotenv.2023.165704

Zubkova, M., Lötter, M., Bronkhorst, F., & Giglio, L. (2024). Assessment of the effectiveness of coarse resolution fire products in monitoring long-term changes in fire regime within protected areas in South Africa. International Journal of Applied Earth Observation and Geoinformation, 132, 104064. https://doi.org/10.1016/j.jag.2024.104064