A Study on Wildfire Behavior Prediction Using a CNN-LSTM Model

Melissa Lin, Rocco Pieczonka, Sai Gayathri Muthukumar, Troy Eigenmann

April 2025

Abstract

Wildfires remain one of the most prominent natural disasters in the United States. Every year the U.S. government spends billions of dollars mitigating their spread. Unfortunately, there are not many softwares available to predict the spread of wildfires. The few software tools that are available to firefighting agencies are often computationally inefficient or narrow in scope. In order to remedy this, we attempted to create a convolutional neural network (CNN) trained on satellite data capable of rapidly predicting the spread of wildfires anywhere in the U.S.

The data used to train this model includes infrared and RGB photos of wildfires as well as weather data from the region. The images came from the Sentinel satellite missions via Sentinel Hub while the weather data was acquired from NASA's POWER API. Two main models were created, the first was trained on data from a single fire in Hawaii which lasted an entire year, while the second was trained on over 5000 short-lived wildfires located all across the country. The first model initially struggled to predict wildfire spread but was significantly improved after introducing bias, achieving a final MAE of 64.37 percent. However, the model proved incapable of predicting out of sample wildfires. In response to this, we created the second model. While better at predicting fires outside of the training set, this model performed at an unsatisfactory level, with an MAE of 19.60 percent.

Despite poor accuracy, our CNN showed promise in assisting wildfire agencies in their battle against wildfires. Future work can be done to make this model significantly more accurate. The next step would either be to downgrade the spatial resolution of the wildfire images, or pay to access geospatial data such as elevation at a much finer spatial resolution.

1 Introduction and Background

1.1 Introduction

Climate change has had an instrumental effect on the severity of natural disasters, which has become apparent in the past twenty years [35]. One of these natural disasters that has become much worse and much more common is wildfires [18]. Wildfires can be extremely devastating to both the environment and humans, displacing both, and causing billions in damage every year [33]. But most tragic, is the loss of human life that occurs because of wildfires. This is why the government has been trying to track, monitor, and stop wildfires since as early as 1885 [5]. Predictably, as time passed, our systems for detecting, monitoring, and fighting wildfires have changed dramatically. While it started as having forest rangers lookout for fires, it has evolved to using satellites to remotely detect wildfires. However, even newer is the process of wildfire behavior monitoring.

The implications of remote satellite wildfire detection are numerous, from identifying areas that need to be evacuated first to identifying where firefighting resources should be directed. Such systems are vital to saving as many human lives as possible. However, it is also critical to understand that not all wildfires are bad. The US Government has previously tried to stop all wildfires and that ended up making wildfires worse

[7]. Wildfires are always going to happen and some of them are critical for preventing larger more extreme wildfires. Yet, we still need to identify and monitor them to prevent people from being put in harms way. This is why having resources for identifying and tracking wildfires at any given point is so important; human lives are on the line, and these people deserve to be protected.

1.2 Background

Wildfires are amongst the most damaging natural disasters that affect our ecosystem tremendously—and they are on the rise due to climate change. Satellite imagery, combined with advanced artificial intelligence techniques such as machine learning and data mining, has proven to be a powerful tool for predicting wildfires. Through the usage of remote sensing data, researchers can gain valuable insights into environmental indicators like vegetation health (NDVI), land surface temperature (LST), and thermal anomalies to forecast fire risks with high accuracy. The study utilized big data platforms like Databricks to process the data, train the model, and validate the approach using multiple classification metrics, cross-validation techniques, and comparisons with existing wildfire early warning systems [30].

Usage of machine learning for wildfire science and management has increased since the 1990s. Applications include fire detection, risk assessment, climate impact analysis, and fire behavior prediction. In older publications and research, traditional models such as random forest, support vector machines, and neural networks were employed, but there is a growing demand for the integration of more advanced methods like deep learning for analyzing large-scale spatiotemporal datasets. As wildfire prediction relies heavily on identifying patterns across space and time, deep learning approaches like recurrent neural networks (RNN) and long short-term memory (LSTM) models have been explored. A study that considered two sets of remote-sensing data—Landsat-8 and MODIS—developed a wildfire susceptibility analysis integrated with deep learning models, namely RNN and LSTM. The results showed that temperature, wind speed, and slope are crucial environmental factors that lead to wildfires, and RNN assured better prediction accuracy than LSTM [15][3].

Remote sensing data for this purpose comes primarily from satellite imaging, as mentioned above Landsat-S and MODIS. MODIS (Moderate Resolution Imaging Spectroradiometer) is a key instrument on NASA's Terra and Aqua Satellites [19]. Landsat-S is another set of satellites that have been collecting images of the earth since 1972 [10]. These are two critical satellite constellations that have been used across various studies for wildfire identification, however the Sentinel Missions are also critical to the majority of studies that have attempted to track and identify wildfires. The Sentinel-1 Mission, the Sentinel-2 Mission, and the Sentinel-3 Mission are the most cited and are the ones we will focus on. Each mission consists of two satellites and has polar orbits [27, 31, 8]. They typically have a repeat time of twelve days for any location [27]. Each mission has a different primary instrument that allows them to capture different aspects of the earth. It is also important to know that these instruments capture different spectral bands. Spectral bands being the different wavelengths a satellite sensor can pick up. The Sentinel-1 satellites primarily use a C-SAR (C-band Synthetic Aperture Radar) instrument, which picks up a wavelength of 1530-1565 nm [27]. The Sentinel-2 satellites have a Multispectral Imager (MSI) instrument as one of its key instruments, which captures high resolution images of the earth's surface [31]. The sentinel-3 mission has a key focus on helping identify climate change, as such it has a Sea and Land Surface Temperature Radiometer (SLSTR) which is critical in wildfire identification because we can detect temperature anomalies using it [8]. These satellites are instrumental to wildfire identification and will likely remain the prominent sources of data for the foreseeable future.

Much of the existing research focuses on wildfire detection and assessing fire severity in real time. Many studies aim to improve fire detection methods using satellite imagery, thermal anomaly identification, and deep learning models that classify burned areas. Besides, machine learning models have been applied to assess the intensity of fire by analyzing temperature anomalies, fuel availability, and environmental conditions. However, while the detection of wildfires and classification of their severity have seen significant interest, less emphasis has been placed on predicting the spread of a wildfire. Understanding where a wildfire will move next is crucial for effective mitigation strategies, resource allocation, and evacuation planning.

This study tries to fill this knowledge gap by proposing a wildfire trajectory prediction model, which forecasts the geographical spread and magnitude of the wildfires. Other than typical fire detection models, which identify the presence and severity of fire, this approach will make use of deep learning to analyze spatiotemporal patterns in wildfire progression. It will also try to predict the direction, speed, and intensity of movement of wildfires in real time, by incorporating data from remote sensing and meteorological factors. This is very important for the emergency response teams and policymakers to take proactive measures in fire-prone regions.

In the future, satellite imagery wildfire prediction models will be improved with meteorological parameters such as air temperature, wind conditions, and soil moisture. These will further advance the fire risk assessments and early warning systems, thus potentially reducing damages due to wildfires. However, the capability for modeling in wildfire spread is an underexplored research area. The integration of satellite imagery with deep learning-based trajectory forecasting could provide a novel solution to effectively predict the movement of wildfires and ultimately assist emergency responders in proactively managing fire outbreaks. [30].

1.3 Methodologies in Literature

Wildfire spread prediction is vital for planning fire suppression and evacuations ahead of time [15]. Unfortunately, physical and empirical simulations modeling wildfire spread are computationally expensive, which makes them difficult to use in influencing firefighting decisions. However, machine learning models are also capable of wildfire spread prediction and are much more computationally efficient [2]. This allows them to produce predictions at a much faster rate which opens the door for their use in firefighting management decisions.

The earliest use of machine learning in wildfire spread prediction was in 2004 when Vakalis et al utilized a radial basis function (RBF) neural network in order to optimize a fuzzy logic system that modeled fire spread in a mountainous Greek forest. Optimizing the logic system with the RBF neural network allowed the system to be time efficient while maintaining wildfire prediction accuracy. Overall, the neural network brought the computation time of the fuzzy logic model down an entire magnitude, allowing its predictions to be used in real time by firefighting agencies [34].

A few years later in 2009, Markuzon and Kolitz were the first to utilize machine learning entirely to predict whether a wildfire would grow or not. They chose to use random forest, Bayesian networks, and k-nearest neighbors, which they trained using remote sensing data from the MODIS and Landsat satellites combined with NOAA data from local weather stations. No model significantly outperformed another. Despite decent accuracy scores, these models did produce a large amount of false positives which makes it difficult to recommend their use in wildfire management [25]. Similarly, a study done by Kozik et al in 2013 also utilized remote sensing data and machine learning models to predict wildfires. They trained a recurring neural network (RNN) acting like neurons in a grid to predict fire spread and optimized it with a Kalman filtration RNN. The optimization RNN improved accuracy but became time consuming when predicting large wildfires [21]. A year later, Kozik et al improved upon this model by introducing wind with both speed and direction as well as the ability to test the effectiveness of firefighting measures in real time [22].

In 2017, Subramanian and Crowley utilized satellite images to train reinforcement learning models to predict the spread of two major wildfires in Canada. Value iteration (VI) and asynchronous advantage actor critic (A3C) were the models trained. These models were given fire ignition point on a grid and told the fire can go in any of the four cardinal directions. A reward function was then defined based on whether a cell was burned or not. Both models were able to achieve a strong accuracy score, but A3C performed slightly better with an accuracy score of about 83 percent. The main two limitations of these models are their inability to be used for real time firefighting decisions due to their long computation time, and their relatively few environmental inputs. For example, they completely ignored wind speed and direction [32]. A year later, Subramanian and Crowley collaberated with Ganapathi to further test reinforcement learning models on wildfire spread prediction. Ultimately, asynchronous advantage actor critic remained the best model, this time with an improved accuracy score of about 87 percent due to improved environmental inputs [9].

In 2017, Zheng et al used a cellular automaton (CA) integrated with an extreme learning machine (ELM) to simulate forest fire spread. They trained their model using five wildfires that occurred in the United States. Unlike most models mentioned already, the ELM excels at handling the effects that wind has on wildfire spread. The integration of the ELM caused the model to outperform traditional CAs with an accuracy of around 80 percent. However the model struggles to deal with rivers and other fire barriers [37].

Back to neural networks, in 2019, Hodges and Lattimer utilized a deep convolutional inverse graphics network (DCIGN) to predict wildfire spread. Datasets generated by one of the most popular fire simulation models, FARSITE, were used to train the model. This neural network predicted wildfire spread in increments of 6 hours, up to 24. They were successful and achieved a precision of .97 and an F measure of .93. This is highly accurate and the model runs two magnitudes faster than the FARSITE simulation itself. However, the model keeps the weather static which limits its potential. [12].

In the same year, Radke et al. created another convolutional neural network they coined as FireCast. This model was trained using geospatial satellite datasets as well as weather datasets from the NOAA. Interestingly, they used the normalized difference vegetation index (NDVI) to assist in the training of this model, which is a measurement of vegetation health calculated using different spectral band photos taken by satellites. The end result of this model shows an accuracy rating of about 88 percent. This is a much better score than the FARSITE simulation was able to produce, but slightly lower than Hodges and Lattimer's DCIGN neural network of the same year. Some limitations of the model include the fact that it was only trained using data from the Rocky Mountains, and the poor spatial resolution of the weather data [28]. A year later, Zhai et al also used a neural network, specifically a multilayer perceptron (MLP), to map wildfire spread. However, they were more focused on short term speed and direction of fire spread rather than long term location prediction. This model was trained using a combination of simulated data and controlled burn data and was found to be highly accurate. It was even able to closely map the spread and speed of a fire in a controlled shrub land environment. The biggest two limitations are its lack of generalizability and inability to give long term predictions [36].

Another study was conducted using neural networks in 2022 by Allaire et al. In this study, they created a deep neural network (DNN) with the goal to replace a computationally expensive fire simulator called ForeFire. Datasets for model training were generated by ForeFire, which simulated 10,000 fires in the Corsica Island region. The DNN they created ran 15,000 times faster than the ForeFire simulation itself and got very close to the simulation's accuracy, making it very useful for real world firefighting decisions that require real time predictions. The biggest limitation of this model was its lack of dynamic wind; the model maintains static wind speed [2].

Stepping away from neural networks, Cheng et al (2022) chose to use a hybrid machine learning approach to predict wildfire spread. This involved using the reduced order modeling techniques: principal component analysis (PCA), convolutional autoencoder (CAE), and single value decomposition autoencoder (SVD AE) as well as random forest (RF), K-Nearest Neighbors (KNN), and Multi Layer Perception (MLP). Datasets were taken from real fires in California as well as generated using a CA fire simulator. They achieved their goal of increasing the computational speed while maintaining accurate prediction scores. The RF and MLP models achieved accuracy scores within 5 percent of the simulation scores, while the CAE was more generalizable but took longer to compute.[6].

Additionally, Radocaj et al (2022) chose to analyze similar models including RF and KNN. However, they chose to use data from LANDSAT and MODIS satellites to train their models instead of simulation data. The models produced wildfire spread predictions in hourly intervals. Results showed that RFs using all eight LANDSAT bands as input produced the highest two day model accuracy at 91 percent. Compared to previously mentioned machine learning models, this is by far the highest accuracy achieved for a prediction this long after ignition [29].

In 2023, Jiang et al created a hierarchical convoluted neural network which is unique in the fact that it utilizes a spread spatiotemporal distribution field which measures fire spread as a function of time and space. To train the model, they used a real world fire from 2019 in California as well as a simulation ran by FARSITE. The model predicts the fire spread up to 12 hours after ignition. Similar to the other neural networks here, the model was successful in getting similar accuracy scores to the simulations while taking a fraction of the time to run [17].

Another neural network was created in the same year by Marjani et al which they called FirePred. What separates it from other neural networks is that it is a hybrid neural network. They combined multiple CNNs and recurrent neural networks (RNN) which allowed it to process data at different intervals ranging from real time to daily. Results for this model exceeded previous neural network models by far. They achieved an F1 score of 91 percent while other models including DCIGN sat around 70 percent [24]. Shortly after, Marjani collaborated with Mesgari to test multi kernel CNNs against normal CNNs and found that using multiple kernels significantly boosts CNN performance. They managed to increase a basic CNNs testing F1 score of 58 percent to 71 percent [23].

Datasets are often a limiting factor when it comes to wildfire machine learning predictions. Recognizing this, Huot et al created a dataset called Next Day Wildfire Spread. This dataset contains almost a decade of remote sensing data in the United States and contains 18545 wildfires. Each fire in the dataset is represented by two data points, one at ignition, and one a day after. The main limitation of this dataset is the poor spatial and temporal resolution: the spatial resolution is 1 kilometer, while the temporal resolution is a day [14].

2 Methods

2.1 Data Preparation and Preprocessing

The first of many steps we had to take to prepare our data was simply selecting a subset of data to initially build our model. Since we have a large quantity of data to work on, we thought it best to choose a small subset of wildfires and their data for building our initial model. This way we will be able to quickly test and develop our model. We plan to build a system that is trained on one wildfire's behavior at a time. As such we decided to pick only one fire to start building our model on, and then expand the model to account for multiple wildfires. We chose a fire that happened in 2022 on a small mountain in Hawaii, it lasted for nearly the entire year which meant there was an abundance of data for this fire that we can and will use.

Now with our sample wildfire selected we need to build two preprocessing systems, one that can calculate the Normalized Difference Vegetation Index (NDVI) and one for calculating the Normalized Burn Ratio (NBR). Both of these indices are vital to accurately predicting wildfire behavior because they tell us about the area surrounding the wildfire. The NDVI tells us how wet and dense the wildlife is around the fire, and the NBR tells us what was already burned. Both of these can be calculated directly from infrared and RGB satellite images.

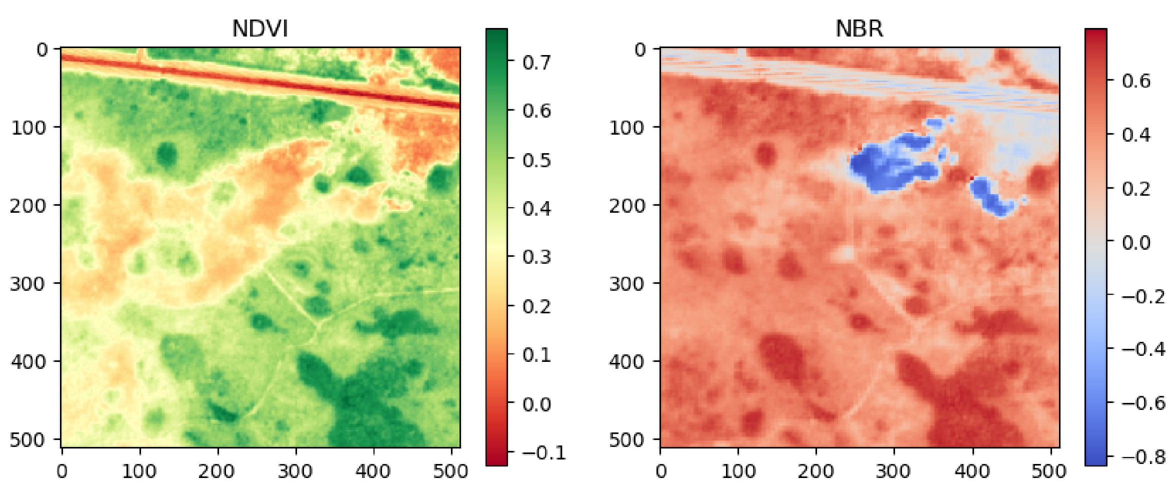

Figure 1: Heatmap of NDVI and NBR, where for NDVI Green is Healthy Vegetation, and for NBR Dark Blue is Burnt Areas

The NDVI and NBR can be calculated from the bands contained within Infrared images. These bands are B12 Shortwave-Infrared (SWIR), B08 Near-Infrared (NIR), and B04 Red (RED). We can pull individual bands from our infrared .png images using a simple normalization process. Then we can calculate our NDVI and NBR using these formulas [38, 1]:

NDVI = $\frac{NIR-RED}{NIR+RED}$, and NBR = $\frac{NIR-SWIR}{NIR+SWIR}$

Using these formulas we can apply our NDVI and NBR to all of our infrared images collected for our satellite images. This process is quick even at a large scale. We then saved this data as .npy files and stored them in a folder adjacent to our wildfire satellite images.

2.2 Edge Detection Model Selection

To prepare the raw images for predictive modeling, we used edge detection techniques. We used different techniques such as Sobel, Canny, and Laplacian to determine the best method to extract fire boundaries. We selected the Canny Edge Detection algorithm as the best method due to its ability to capture finely grained contours and reduce noise. To improve its accuracy, we also included preprocessing steps which included grayscale conversion, Gaussian blurring for further noise reduction, and adaptive thresholding (Otsu's method) to enhance the actual perimeters of the fires, fire-affected regions, and the background.

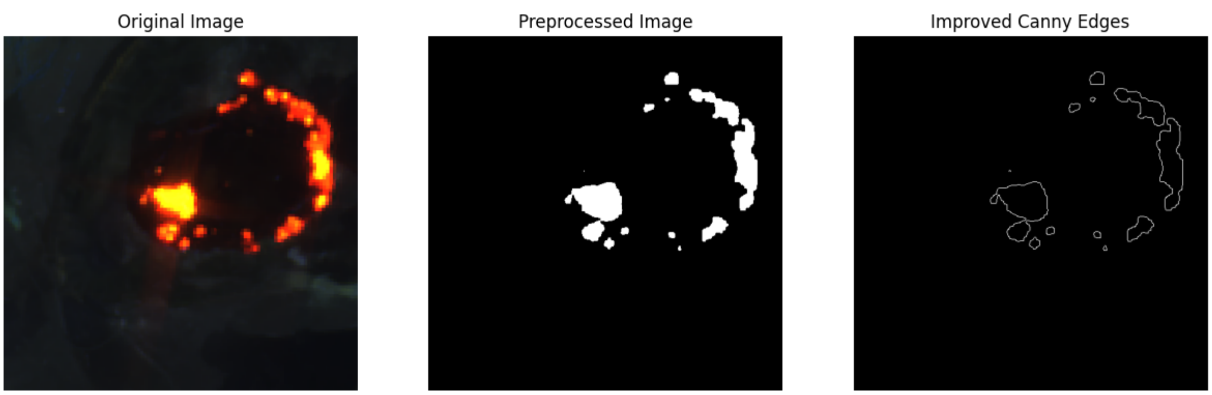

Figure 2: The original image (left) displays raw satellite data. The preprocessed image (center) is a binarized version from adaptive thresholding, isolating fire affected areas. The improved canny edges (right) shows the extracted contours.

2.3 Weather and Elevation Integration

The weather data has already been collected and, as such, have no necessary preprocessing. The notable features of weather data are wind speed, direction, and precipitation (includes rain and snow). However, as has been seen in our exploratory data analysis, most of the wildfires in our data are in the lower United States, meaning it is unlikely that we have any records of precipitation that are snowfall. As for elevation data, it was collected via the coordinates of the wildfire that we are currently working with is 900 meters above sea level. This was determined manually, but in the future we will collect this data using an API similar to the weather data and image data.

Now in terms of integration, weather data and elevation data will be included in our RNN model since RNNs handle numerical data better than CNNs. This is the main reason why we believe a hybrid neural network will be the optimal model to work with since it can handle images in the CNN and text data in the RNN and then determine patterns between the two models.

2.4 CNN and RNN Integration and Implementation

Convolutional Neural Networks (CNNs) are a fitting choice of model to predict wildfire behavior as they are well-suited for image-based analysis. Our wildfire behavior prediction model leverages this to extract valuable spatial features from our Infrared and RGB satellite images to compute the vegetation values like NDVI and NBR above. The CNN architecture is structured to detect spatial patterns such as fire intensity, vegetation density, and burnt areas. We can also use operations such as MaxPooling in order to reduce dimensions while not compromising on the retention of key features. Key features can be flattened and passed through layers to make accurate predictions on fire probabilities and intensity. Activation functions we can use include Rectified Linear Unit (ReLU) to fully connect layers and improve feature learning and Softmax/Sigmoid for classification tasks such as identifying burnt vs unburnt areas.

We can train the CNN based on our historical satellite imagery which is organized by the identifier column that includes longitude, latitutde, and date in yyyymmdd format. The model will learn to identify the areas that are more prone to fire and develop a probability in a given area. It will also be able to detect active fires and classify various land conditions that aid model forecasting efforts.

The implementation of the CNN is rather straight forward. We will split the data for each wildfire into an ≈80-20 train-test split. I say approximately because some (short) wildfires will not have enough data (minimum 5 images) to be properly split this way. If that is the case we will adapt the split accordingly. There is a strong chance that we will not be able to train our model on a wildfire's behavior if it does not have at least 3 images of the fire; however, this is not yet confirmed. We plan to experiment with different activation functions and number of layers in the CNN. As for batch size and epochs, these will also be tested with different numbers to determine what produces the best results.

Since CNNs are best used with image data, we will need to implement a different type of neural network within our model to accommodate our text-based data. This includes weather data, elevation data, and any other text data we deem beneficial to our model's performance. The current plan is to utilize recurrent neural networks (RNNs) to handle and process this text data. RNNs were chosen as they showed promise in past research in wildfire prediction using machine learning models [20]. In general, they are best known for their ability to utilize previous outputs as inputs. This may prove useful as some of our wildfire photos are entirely obscured by clouds, which will leave gaps in our wildfire timeline.

Recurrent neural networks have multiple parameters that can be adjusted which will allow us to tune their performance. Our focus will on testing differing amounts of recurrent layers, adjusting the amount of neurons per layer, as well as changing the activation function in each layer. Since the amount of parameter combinations is vast, it is unlikely that we will find the perfect model without significant time investment. However, after a while it should be obvious as to whether or not the RNN is going to be up to standard. If it is not, we can look to utilize alternative neural networks.

In order to utilize the outputs from both neural networks, a merge layer will be used. This way we can build two separate models and then merge their outputs before passing it through a final dense layer. This should allow us to make a nice segmented design that is easy to code, troubleshoot, and tune. The RNN model will be built similar to the CNN model, except of course, its input will be our numeric data.

Our plan for building this model is to start by making a simple CNN model and then expanding it to incorporate more features and then do the same for the RNN. Then we will merge them once they are individually complete, and make a final layer (or few layers) to integrate their results. The reason we plan to take a progressive approach is to ensure that each step of the process is accomplished effectively when building our model. This model is going to be quite complex and as such building parts is easier than building it all at once. This also helps us to ensure that each part of the model improves its accuracy. This approach also helps us to divide the tasks more easily among ourselves. Additionally, this method will allow us to spend time at each step to examine patterns and trends among our results, potentially leading us to some interesting conclusions.

3 Results

Before diving straight into the results and implications of our model we first should mention a few things. First, the model was initially constructed using only one wildfire. This was done to make the development and testing process quicker and easier. It is currently being expanded to a larger set of wildfires. The second item to mention is the variance in the data. Wildfires are often quick and destructive; as such the data that we have is not always well suited for training the model. Our current models take around 3-4 input frames to capture trends over time. But the satellites we have often only get a single picture of a location every five days. Meaning those 3-4 input frames require 15-20 days to collect (most of the time) [13]. After 20 days most wildfires are contained or start dying out. Unfortunately there is not much we can do about this flaw in the data. Keeping this in mind we decided to build the model so that it could take 3-4 input frames over any reasonable time frame, be it hours or days, and then make predictions based on that. Knowing all of this will make the model development more understandable.

3.1 CNN-LSTM Architecture

To begin let us dive into the model architecture we currently have created and break down its layers, then lets dive into the results it gave us.

The current CNN-LSTM model has 14 layers, with the first three layers are all about the weather data. The first layer is an input layer where the weather data gets taken in and 5 time steps are created for 5 frames of the weather data. Each frame contains the weather features of precipitation, wind speed, and wind direction, which is crucial for contextualizing the environment in which the fire occurred. The second layer applies a dense layer to each of the five time steps to understand complex connections between weather patterns. This is done using the TimeDistributed Dense layer which transforms each 3-feature vector into a 32-dimensional feature space which is used to investigate the dynamics of how low precipitation and/or high wind speed may be indicators of fire. The third layer is a dropout layer that randomly drops 20% of the connection values to prevent overfitting. Specifically, the model will be able to better generalize and not rely too heavily on specific features of the training fire.

From here on the layers are more integrated and complex. The fourth layer is a simple input layer for the Satellite images and it splits them into 5 time steps (one time step is one satellite image frame). Each frame contains the NDVI and NBR channels to provide insight on the vegetation and burn severity over time. The fifth layer is a LSTM layer specifically for the weather data to capture temporal patterns in the weather data, such as sustained dryness or shifting wind over a period of time. The sixth layer is our first CNN-LSTM layer, it combines CNN and LSTM operations to process the NDVI/NBR image sequence (where a sequence is 5 time steps). Now the weather data is a simple csv input and so it has to be expanded to work with our image data. In the seventh layer the weather data (precipitation, wind speed, and wind direction) are each mapped to a 512 x 512 grid, where each pixel in the grid is labeled with the precipitation, wind speed and wind direction. This allows the weather data to be directly placed on top of the NBR/NDVI like a simple 4D tensor (x, y, z, time, where z is the NDVI stacked on NBR stacked on the weather data).

From here our data is fully integrated and ready for true ML. The eighth layer is another CNN-LSTM layer which takes the complex inputs up to this point and outputs a single 512 x 512 x 64 feature map which describes how the fire spreads over the course of the sequence of 5 time steps. Layer nine reshapes the output to be 512 x 512 x 8, to make it integrate well with the weather data. Layer 10 is a normalization layer and ensures the image features are normalized, making training more stable and helping the model converge faster. The eleventh layer is a convolutional filter applied to the weather data to refine its spatial features. The twelfth layer finally concatenates the weather data and the satellite image data, producing a result that combines the spatio-temporal patterns from the images with the weather effects. The 13th layer is another convolutional layer that learns interactions between image and weather features over a spatial area. The final layer is the output layer which produces the final predicted NDVI/NBR map for the next timestep. This last layer also includes a bias term. Since the wildfires generally take up less than 10% of the actual image, the model would naturally predict no fire on the entire model and generally it would be fairly accurate. However, if we add a bias term we can curb this issue by making the model more prone to predict a tile as on fire than it otherwise would.

The bias term in our model is one of the most critical components to refine. Shown below is a figure of the model without the bias term and one with the bias term. These were our preliminary results.

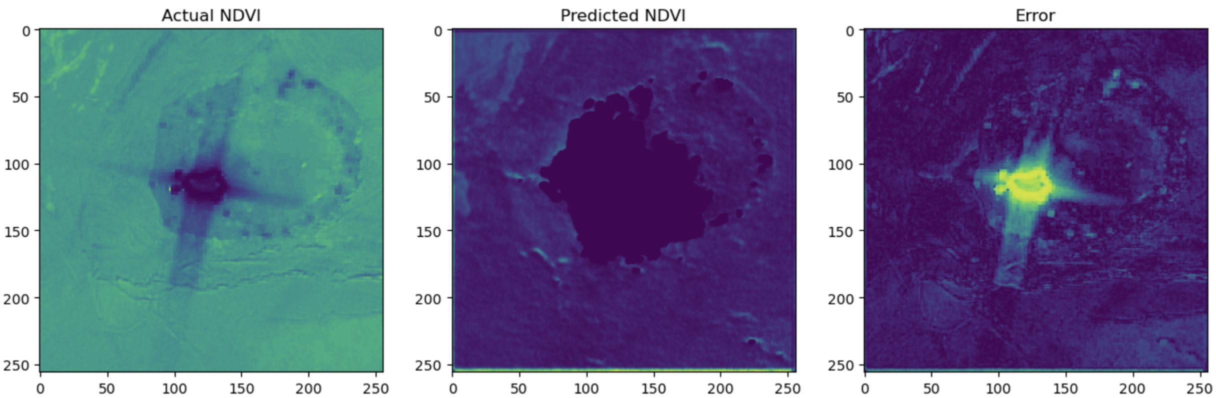

Figure 3: NDVI Actual vs Predicted vs Error without Bias

The MAE of this model was 91.12% which seems great on paper, but one look at the actual results and we can see that this number is useless.

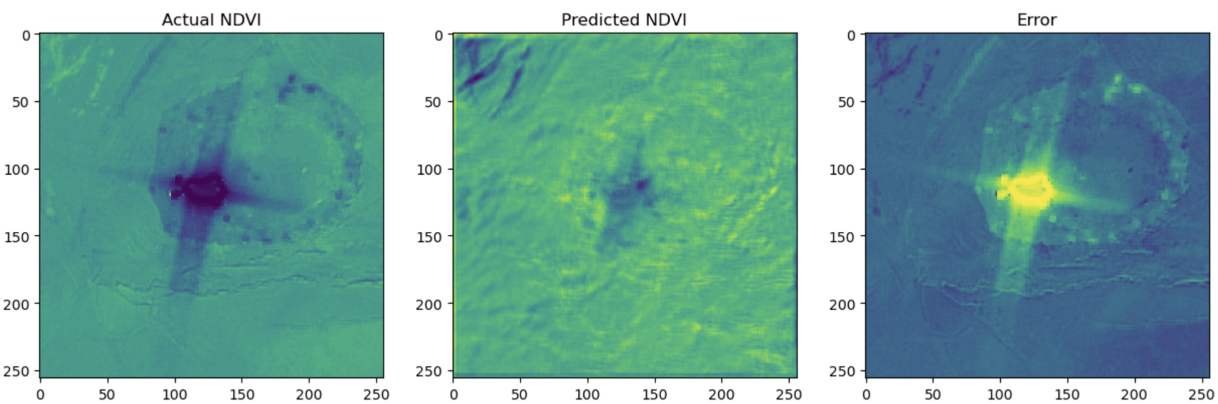

Figure 4: NDVI Actual vs Predicted vs Error with Bias

The MAE of this model was 64.37% which is significantly less than the previous model, but here we can see a vague outline of the wildfire. This gives a much more useful idea of how the fire will spread over time, however it is not perfect and will be refined more in the Results section.

3.2 CNN-LSTM Results and Implications

With the model's architecture fully described, let us dive into the results we were able to get with this model. First and foremost we set the batch size to 4 and the number of epochs equal to 1. Unfortunately increasing the epochs only leads to over fitting every time, so for now the epochs are kept to one.

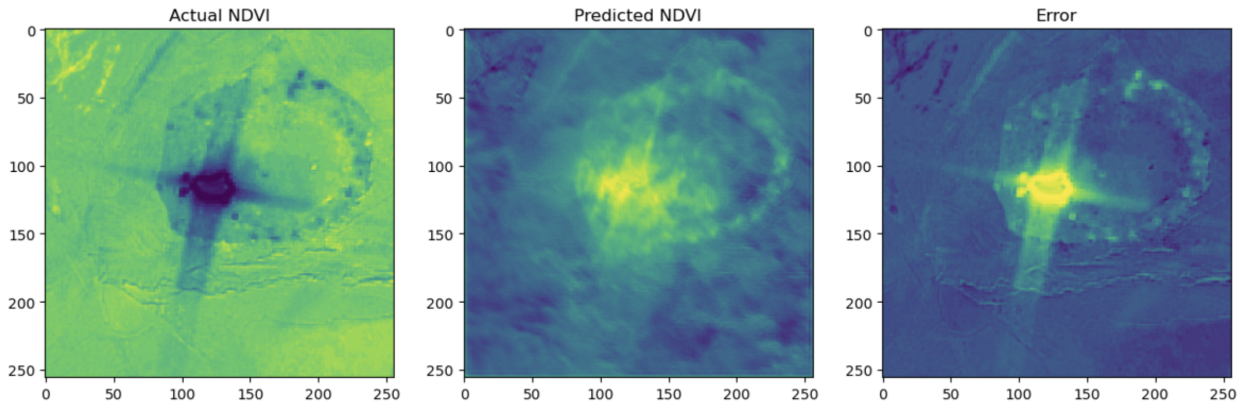

Figure 5: NDVI Actual vs Predicted vs Error

In the example shown above the MAE is 0.6730, which is not great, but it does decently capture the fire spread. Additionally in the above figure, the color scheme for Actual NDVI and Predicted NDVI are flipped, if you invert the colors for Predicted NDVI you find that the images are fairly similar in terms of fire prediction.

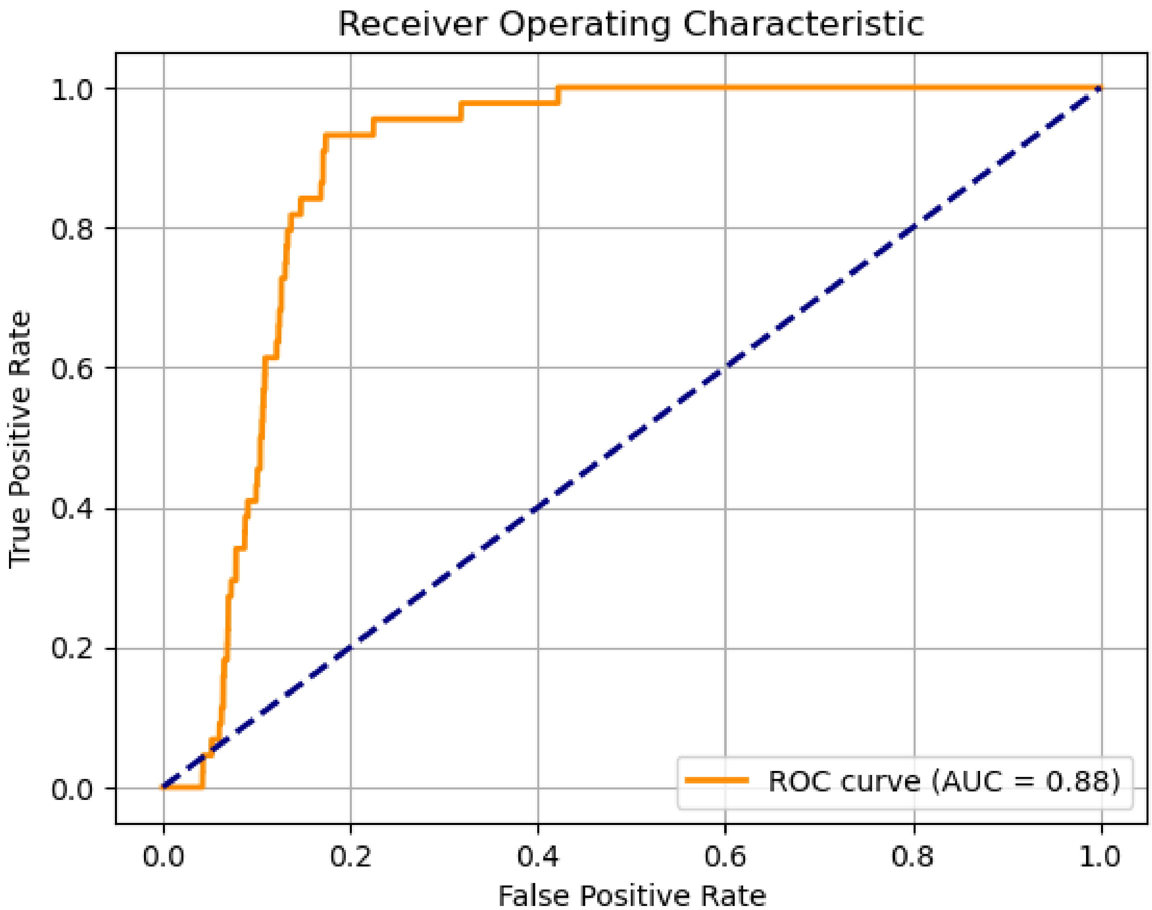

Figure 6: The ROC curve shows the model's performance in detecting fire-affected areas using the predicted Normalized Burn Ratio (NBR) values from the fire channel (channel 1). The AUC score of 0.88 shows a strong discriminative ability in identifying burned regions.

The curve here is incredibly promising because it shows that the current model is better than random guessing. While not perfect by any means we do not expect this model to reach insane levels of accuracy because wildfires often do not behave as expected. Our AUC however, tells us that this model is reaching good levels of accuracy.

The implications of our model show that there is merit to a CNN-LSTM based wildfire prediction model. Although ours is far from complete, it demonstrates that there is great potential for the model and emphasizes the need for more testing on larger, well-constructed, datasets.

3.3 CNN-LSTM Trained on Multiple Fires

After getting the results on the previous model, we decided to re-train the model using twenty wildfires instead of the one Hawaii wildfire. Instead of using hundreds of photos throughout a single wildfire's lifespan, we chose to use two photos per wildfire. One during the wildfire, preferably towards the ignition time, and one after photo. Unfortunately, due to the poor temporal resolution of the Sentinel 2 satellites, some of the after photos were taken once the fire had already extinguished. However this was deemed to be acceptable as the burned areas are clearly visible in the NBR and NDVI images.

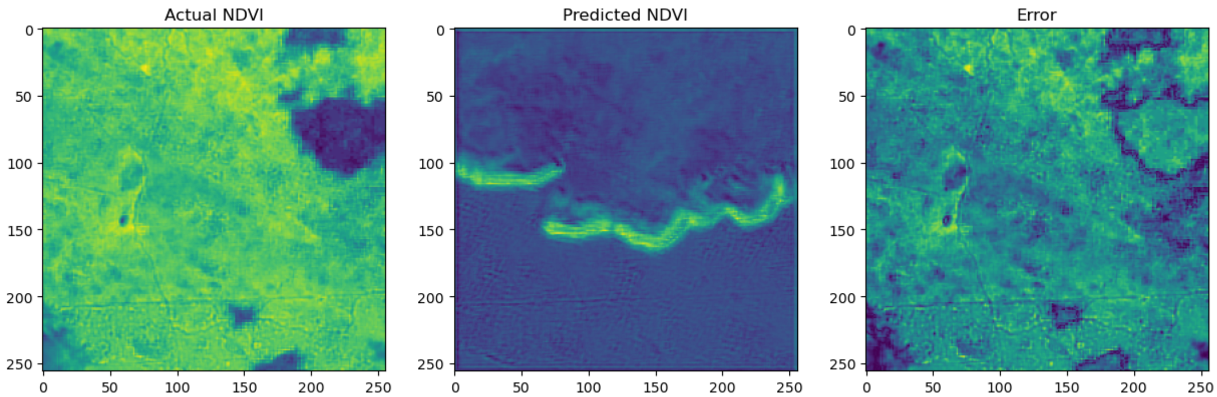

Figure 7: NDVI Actual vs Predicted vs Error from the Model Trained on Multiple Fires

Although the MAE was significant at an average of .7, it is apparent that this model significantly underperformed compared to the previous model. The predicted NDVI looks nothing like the actual NDVI. In fact, the prediction still shows the same fire waves that were in the before image, which suggests that this model is not doing much predicting. This is most likely because the dataset the model was trained on only contains twenty fires, which is simply not enough data for the model to learn the complexities of wildfire behavior on.

3.4 CNN-LSTM Trained on all Fires

Using the basis of the previous model, we expanded it to all of the wildfire data that we currently have. This data is all of the data from 2022 U.S. wildfires that were identified by NASA's MODIS satellite [26]. We had to collect some more photos from Sentinel Hub to ensure that every wildfire had at least 2 associated images. Once every sequence was established we ran the code and got the results for our final model.

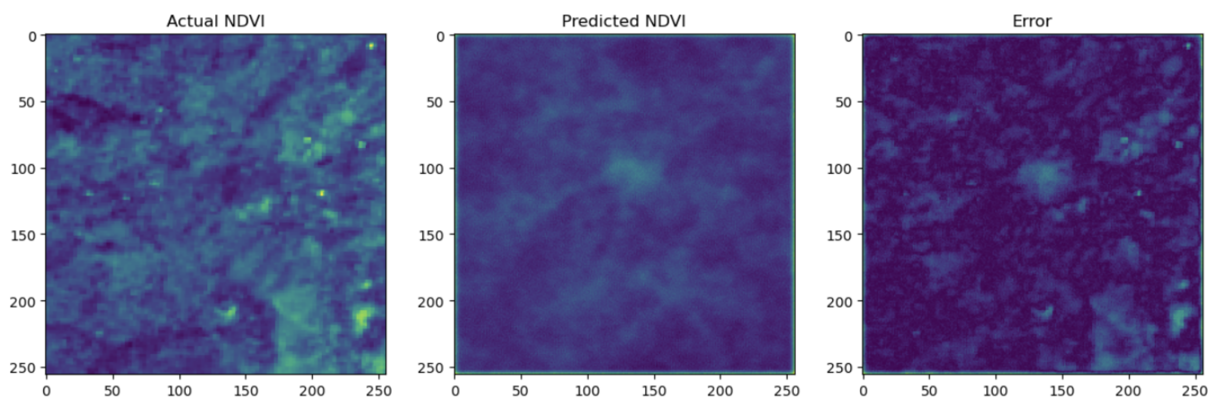

Figure 8: NDVI Actual vs Predicted vs Error from the Model Trained on All Fires

These results were quite disappointing. Our bias variable was tuned to the one wildfire that we used to build our model. As such it seems that that bias value did not translate well to our final model. This model did produce a MAE of .1960, which appears good, however likely means that our bias variable needs to be better tuned. Perhaps as well the number of filters and layers in our model would also need to be tuned.

For additional reference here is an image of the actual vs predicted NBR, which ends up being quite a bit worse than NDVI

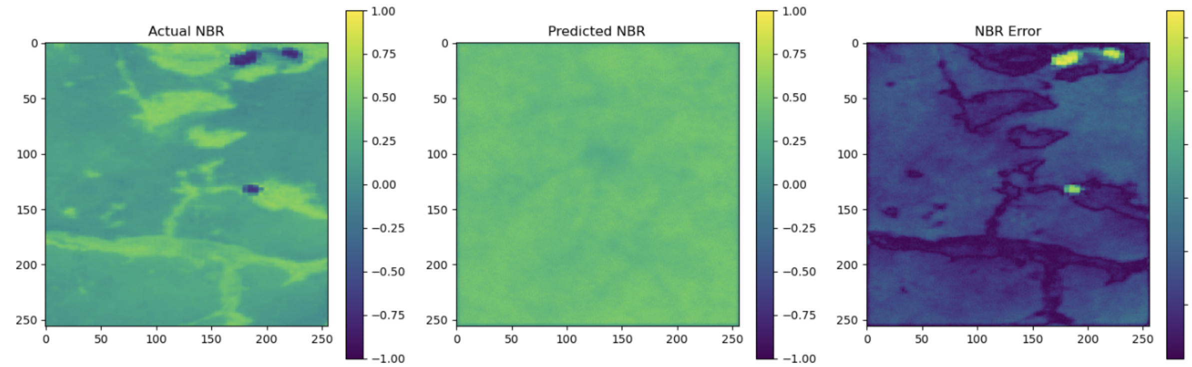

Figure 9: NDVI Actual vs Predicted vs Error from the Model Trained on All Fires

It seems that the model we are building does not understand complex wildfire spread patterns. Perhaps this is due to the uniqueness to each wildfire, or maybe there are too few inputs for the model to properly understand wildfire spread. While a well tuned model works for an individual wildfire, it seems that it currently does not expand well when learning on multiple different types of wildfires.

4 Discussion

4.1 Hawaii Model

The Hawaii Model demonstrates decent integration of the temporal and environmental data for wildfire prediction. The incorporation of the convolutional LSTM layers allow the model to more effectly capture the spatial-temporal dependencies in the NDVI and NBR imagery, while also incorporating the weather data to provide context. The multi-input approach proves to be useful to handle the non linear dynamics of wildfires. The results, including an AUC score of 0.88 and MAE of 0.67, suggest that the model can identify the burn areas well, although the result is blurry and unclear for human interpretation.

Overfitting was a challenge encountered throughout the process of the developing the model. Increasing the number of epochs only decreased accuracy which emphasizes the model's reliance on the early stopping. We used a focal loss function to handle the imbalance between the burned and unburned areas. Since most of the satellite images show small percentages of areas with fire, using the focal loss helped the model better identify the areas that we are looking to predict off of.

Overall, the model set up with the weather data integration can be effective for predicitng wildfires and the next step was to train the model on more fires, which will be discussed in the following All Fires Model.

4.2 All Fires Model

The All Fires Model uses historical wildfire data in attempt to fit the model behavior prediction to the actual data. While the model performed reasonably well in the Hawaii model, combining all fires– human and lightning-caused fires– makes it more difficult to generalize. The MAE of 0.1960 indicates that the model is not accurately capturing the spread behavior across different fires. This can likely attributed to the bias parameter which was originally tuned for only one fire that did not carry over well when applied to other fires that had differing shapes, speeds, and environmental conditions.

A key takeaway is that the model struggles to handle the complexities and uniqueness of each individual wildfire. Each fire burns in its own way depending on factors like wind, vegetation, and terrain as our NBR and NDVI indexes showed. It is possible that the number of layers and filters in the model are not sufficient to capture the spatial and temporal patterns that actually drive wildfire spread across different scenarios. Our All Fires Model illustrates that although deep learning can be promising for fire modeling, our current approach requires more flexibility and data diversity.

4.3 Challenges

By far the largest challenge faced was our models' inability to accurately predict wildfire spread on out of sample wildfires. Even the second model, which was created for the purpose of having a better accuracy when predicting on out of sample wildfire spread, failed to accomplish this. This most likely stems from multiple issues with our model, the first of which is the fact that we were unable to acquire many input variables at the spatial resolution of our wildfire images, which was about 20 meters per pixel. The second being that having 65,536 (256*256) pixels per image was most likely too complex for our model to fully understand the patterns of. Finally, the last issue was the lack of temporal high resolutions in the sentinel satellites [13]. More often than not, by the time the satellite got over the wildfire location again, the wildfire had already extinguished. And even if the wildfire was still there once the satellite returned to the wildfire site, that often meant it was such a massive wildfire that the smoke in the air had completely obscured the satellite's vision. This forced us to predict on the burn area of a wildfire rather than the spread of the wildfire itself. We have known from the start that there exists infinite variables that could affect any given wildfire but it is difficult to predict with limited resources and data to match the exact timings of the wildfire data (in our case, the weather data), especially with some fires that have brief intervals.

All of these challenges can be remedied by lowering the spatial resolution of our wildfire images significantly. Previous work uses a spatial resolutions as low as an entire kilometer per pixel, which would drastically lower the complexity of each image. This would also allow us to easily find other input data to complement each image since geospatial data at lower spatial resolutions is widely accessible. And finally, this would allow us to acquire satellite images at a much higher temporal resolution since lower spatial resolution satellites typically have a higher temporal resolution, and vice versa [4].

Another major issue that we identified is with the broad locations of our data. Research backs up the claim that wildfire spread differs based on location [11]. Basically a wildfire in forested northern California spreads differently than in the New Mexico desert. We hoped that this variable could be overcome by our NDVI and NBR metrics, however it struggles to identify a difference between dry short desert shrubs, and tall health evergreen trees. This trend is clear when you compare our Hawaii wildfire model to our multiple fires model (the model with 20 fires from across the nation). The Hawaii wildfire model was our most accurate one and it was localized to one general location where the wildlife follows similar fire spread patterns. However, for our multiple fires model there is a sharp drop in the model's ability to actually predict anything useful or accurate. This suggests that to accomplish our task of making a general wildfire prediction model, we would need to include location as a parameter and have different focal loss weights and bias weights for each different region.

The last challenge that we ran into is computational efficiency. Our goal was to build a model using a CNN-LSTM architecture which is inherently computationally expensive when considering the size of the images that we have used (20 meter resolution). However, to increase the accuracy of our model we would likely need multiple fully connected layers as well as many more layers. Our current model needed supercomputer access and used around 128 GB of memory. This is no small amount of memory, but to build a more effective model with all of the fully connected layers would need significantly more memory and time to run. Thanks to ASU's Sol Supercomputer [16] we were able to train our model as it is, but should we have tried making the model significantly larger, we would have also wanted better data to make the model more likely to have a useful outcome.

5 Conclusion

Ultimately, the results of our model are unclear, however we can determine that the model in combination with our current data [13] is flawed when expanded to all type of fires in all locations across the U.S. The model we have created is great when tuned to a specific wildfire's behavior, but the model is poor when expanded beyond that type of wildfire. There is merit to our model on a small scale for wildfires, yet when expanded to all the wildfire data that we have, the model behaves poorly and the information it provides is useless.

This model is flawed and unclear, but does have useful implications towards further wildfire research. Our failed expanded model, and rather successful small model, imply that a truly useful wildfire prediction system can likely be developed by building many different models all tuned to a specific type of wildfire. Such as one model for wildfires in woodland areas of northern California and another model tuned to wildfires in the southern Arizona desert. However, to confirm such a hypothesis would require further research. Though, what we can say for certain, is that a general wildfire behavior prediction model based on the data we used, across all of the US, and using a CNN-LSTM architecture to model is ineffective.

References

[1] Emanuele Alcaras, Domenica Costantino, Francesca Guastaferro, Claudio Parente, and Massimiliano Pepe. Normalized burn ratio plus (nbr+): a new index for sentinel-2 imagery. *Remote Sensing*, 14(7):1727, 2022.

[2] Frédéric Allaire, Vivien Mallet, and Jean-Baptiste Filippi. Emulation of wildland fire spread simulation using deep learning. *Neural networks*, 141:184–198, 2021.

[3] Nazanin Bahadori, Seyed Vahid Razavi-Termeh, Abolghasem Sadeghi-Niaraki, Khalifa M. Al-Kindi, Tamer Abuhmed, Behrokh Nazeri, and Soo-Mi Choi. Wildfire susceptibility mapping using deep learning algorithms in two satellite imagery dataset. *Forests*, 14(7), 2023.

[4] Edward H Bair, Jeff Dozier, Karl Rittger, Timbo Stillinger, William Kleiber, and Robert E Davis. How do tradeoffs in satellite spatial and temporal resolution impact snow water equivalent reconstruction? *The Cryosphere*, 17(7):2629–2643, 2023.

[5] John W Chambers. The evolution of wildland fire management and policy. *Fire Management Notes*, 48(2):5–8, 1987.

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

[7] Michael P Dombeck, Jack E Williams, and Christopher A Wood. Wildfire policy and public lands: integrating scientific understanding with social concerns across landscapes. *Conservation biology*, 18(4):883–889, 2004.

[8] Craig Donlon, Bruno Berruti, S Mecklenberg, Jens Nieke, Helge Rebhan, Ulf Klein, Alessandra Buongiorno, Constantin Mavrocordatos, Johannes Frerick, Bernd Seitz, et al. The sentinel-3 mission: Overview and status. In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pages 1711–1714. IEEE, 2012.

[9] Sriram Ganapathi Subramanian and Mark Crowley. Using spatial reinforcement learning to build forest wildfire dynamics models from satellite images. *Frontiers in ICT*, 5:6, 2018.

[10] MohammadAli Hemati, Mahdi Hasanlou, Masoud Mahdianpari, and Fariba Mohammadimanesh. A systematic review of landsat data for change detection applications: 50 years of monitoring the earth. *Remote sensing*, 13(15):2869, 2021.

[11] James E Hilton, Claire Miller, Andrew L Sullivan, and C Rucinski. Effects of spatial and temporal variation in environmental conditions on simulation of wildfire spread. *Environmental Modelling & Software*, 67:118–127, 2015.

[12] Jonathan L Hodges and Brian Y Lattimer. Wildland fire spread modeling using convolutional neural networks. *Fire technology*, 55:2115–2142, 2019.

[13] Sentinel Hub. Sentinel hub, 2025. Accessed on Feburary 16, 2025.

[14] Fantine Huot, R Lily Hu, Nita Goyal, Tharun Sankar, Matthias Ihme, and Yi-Fan Chen. Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022.

[15] Piyush Jain, Sean CP Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D Flannigan. A review of machine learning applications in wildfire science and management. *Environmental Reviews*, 28(4):478–505, 2020.

[16] Douglas M Jennewein, Johnathan Lee, Chris Kurtz, William Dizon, Ian Shaeffer, Alan Chapman, Alejandro Chiquete, Josh Burks, Amber Carlson, Natalie Mason, et al. The sol supercomputer at arizona state university. In *Practice and Experience in Advanced Research Computing 2023: Computing for the Common Good*, pages 296–301. 2023.

[17] Wenyu Jiang, Yuming Qiao, Guofeng Su, Xin Li, Qingxiang Meng, Hongying Wu, Wei Quan, Jing Wang, and Fei Wang. Wfnet: A hierarchical convolutional neural network for wildfire spread prediction. *Environmental Modelling & Software*, 170:105841, 2023.

[18] Matthew W Jones, Adam JP Smith, Richard Betts, Josep G Canadell, I Colin Prentice, and Corinne Le Quéré. Climate change increases the risk of wildfires: January 2020. *ScienceBrief*, 2020.

[19] CO Justice, JRG Townshend, EF Vermote, E Masuoka, RE Wolfe, Nazmi Saleous, DP Roy, and JT Morisette. An overview of modis land data processing and product status. *Remote sensing of Environment*, 83(1-2):3–15, 2002.

[20] Nima Khakzad. Modeling wildfire spread in wildland-industrial interfaces using dynamic bayesian network. *Reliability Engineering & System Safety*, 189:165–176, 2019.

[21] VI Kozik, ES Nezhevenko, and AS Feoktistov. Adaptive prediction of forest fire behavior on the basis of recurrent neural networks. *Optoelectronics, Instrumentation and Data Processing*, 49:250–259, 2013.

[22] VI Kozik, ES Nezhevenko, and AS Feoktistov. Studying the method of adaptive prediction of forest fire evolution on the basis of recurrent neural networks. *Optoelectronics, Instrumentation and Data Processing*, 50:395–401, 2014.

[23] M Marjani and MS Mesgari. The large-scale wildfire spread prediction using a multi-kernel convolutional neural network. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10:483–488, 2023.

[24] Mohammad Marjani, Seyed Ali Ahmadi, and Masoud Mahdianpari. Firepred: A hybrid multi-temporal convolutional neural network model for wildfire spread prediction. *Ecological Informatics*, 78:102282, 2023.

[25] Natasha Markuzon and Stephan Kolitz. Data driven approach to estimating fire danger from satellite images and weather information. In *2009 ieee applied imagery pattern recognition workshop (aipr 2009)*, pages 1–7. IEEE, 2009.

[26] NASA LP DAAC. MODIS Collection 6.1 Active Fire and Burned Area Data, 2022. Accessed: 2023-03-16.

[27] Pierre Potin, Betlem Rosich, Patrick Grimont, Nuno Miranda, Ian Shurmer, Alistair O'Connell, Ramon Torres, and Mike Krassenburg. Sentinel-1 mission status. In *Proceedings of EUSAR 2016: 11th European conference on synthetic aperture radar*, pages 1–6. VDE, 2016.

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

[29] Dorijan Radočaj, Mladen Jurišić, and Mateo Gašparović. A wildfire growth prediction and evaluation approach using landsat and modis data. *Journal of environmental management*, 304:114351, 2022.

[30] Younes Oulad Sayad, Hajar Mousannif, and Hassan Al Moatassime. Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Safety Journal*, 104:130–146, 2019.

[31] Francois Spoto, Omar Sy, Paolo Laberinti, Philippe Martimort, Valerie Fernandez, Olivier Colin, Bianca Hoersch, and Aime Meygret. Overview of sentinel-2. In *2012 IEEE international geoscience and remote sensing symposium*, pages 1707–1710. IEEE, 2012.

[32] Sriram Ganapathi Subramanian and Mark Crowley. Learning forest wildfire dynamics from satellite images using reinforcement learning. In *Conference on reinforcement learning and decision making.* Ann Arbor MI, 2017.

[33] Douglas Thomas, David Butry, Stanley Gilbert, David Webb, Juan Fung, et al. The costs and losses of wildfires. *NIST special publication*, 1215(11):1–72, 2017.

[34] D Vakalis, H Sarimveis, C Kiranoudis, A Alexandridis, and G Bafas. A gis based operational system for wildland fire crisis management i. mathematical modelling and simulation. *Applied Mathematical Modelling*, 28(4):389–410, 2004.

[35] Maarten K Van Aalst. The impacts of climate change on the risk of natural disasters. *Disasters*, 30(1):5–18, 2006.

[36] Chunjie Zhai, Siyu Zhang, Zhaolou Cao, and Xinmeng Wang. Learning-based prediction of wildfire spread with real-time rate of spread measurement. *Combustion and Flame*, 215:333–341, 2020.

[37] Zhong Zheng, Wei Huang, Songnian Li, and Yongnian Zeng. Forest fire spread simulating model using cellular automaton with extreme learning machine. *Ecological Modelling*, 348:33–43, 2017.

[38] Xiaobing Zhou, Huade Guan, H Xie, and JL Wilson. Analysis and optimization of ndvi definitions and areal fraction models in remote sensing of vegetation. *International Journal of Remote Sensing*, 30(3):721–751, 2009.

Appendix A

All project code: https://drive.google.com/drive/folders/1U6qFb7SabrCqVKf3VgBQWFchWbkikGkt
All satellite images and data: https://drive.google.com/drive/folders/1vyOG7sVxMloy7WmYmrybRGkormDxxzZB