Catching Fire: Predicting Wildfire Progress with Computer Vision

Jack Michaels
Stanford University
Palo Alto, CA
jackfm@stanford.edu

github.com/jmichaels32/fireprediction

Abstract

With the ongoing climate change crisis causing spikes in wildfires across the globe, mitigating wildfire spread after ignition has become paramount to supporting environmental success and in many cases save lives. While many local governments are adopting policies to prevent wildfires before ignition, it is common for wildfires to start without human involvement. To aid in the mitigation of wildfires, this paper explores computer vision as a solution to predicting next-day fire spread. We use the ***Next Day Wildfire Spread*** *[14] dataset, a feature rich dataset containing satellite images of historical wildfires and their associated explanatory variables (e.g. topography, vegetation, weather, drought index, population density). Our model aims to achieve stable performance in regimes without data when compared to baselines of data availability. Our data restricted model achieves comparable performance to both the baseline and feature rich models, though further computational resources may be required assess the validity of this gap.*

1. Introduction

Within the past century, human industrial activity has sparked the uncontrolled release of carbon dioxide [6] [17], leading to an escalation in the progress of climate change [16]. This escalation has led to a global increase in average ambient temperature [29] [25], increases in the frequency of heatwaves [4], and regional increases in the duration, intensity, and frequency of droughts [3] [2] [31]. Subsequently, the frequency of wildfires have increased by around 22% globally given historical trends [1] with forecasts estimating that the fire season will increase by more than 20 days per year in northern latitudes within the next century [8]. In addition to frequency, the strength of wildfires have been forecasted to increase, with a modelled increase of around 60% in Europe [10] and up to a 400% increase across regions of historically low flammability in Alaska [32]. This increase in wildfire probability poses a significant threat to future anthropogenic activity and environmental prosperity; the 2020 California wildfires burned 1.7 million hectares, caused 33 deaths, $7.1 billion USD in damages caused by CO_2e emmisions alone [15], and countless long lasting respiratory issues caused by a 40% increase in $PM_{2.5}$ concentrations around wildfires [5].

With this increase in wildfire probability and strength, conventional fire management approaches may no longer be effective in the new era. Alternative methods are being pursued where instead of mitigating the spread of already ignited fires, many local governments are taking preventative action against ignition by better regulating fire creation and subsidizing environmental restoration [9] [23]. Though technically effective, these governmental plans have been met with mild backlash, specifically for their insufficient handling of edge cases exemplified by California's lack of policy in regards to the local native populations' habits [22].

In response to the occasional ineffectiveness of governmental regulation and the occasional nature derived, non-anthropogenic ignitions [27], mitigating and limiting the spread of already ignited fires is a vital technology requiring further development. This paper explores deep learning solutions using both feature rich and data limited models with the overall goal of creating a model capable of predicting fire spread in worst-1 situations where we lack detailed features [33]. This data obscurity is overwhelmingly common during deployment, especially during emergencies such as wildfires, showcasing the importance of performance. We predict next-day fire segmentation masks from previous-day fire segmentation masks (and corresponding explanatory features), of which each mask has three distinct classes: -1 (no data), 0 (no fire), and 1 (fire).

2. Literature Review

2.1. Physics Based

A distinguishing set of fire spread prediction models rely on physics based approaches. While massively reliant on

data observability and precision, these models attempt to capture the fundamental processes governing the fire science of heat exchange.

FARSITE (Fire Area Simulator) uses vector propagation techniques on top of existing models of surface fire, crown fire, point-source fire acceleration, spotting, and fuel moisture to produce fire perimeters at specified time intervals [7]. Figure 1 details specific physic based approaches taken by FARSITE to accurately predict fire spread. Under ideal conditions, FARSITE produced reliable fire spreading predictions yet faltered when presented with more extreme situations such as crown fires.

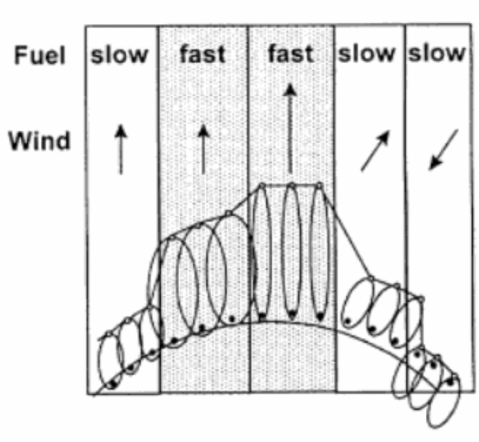

Figure 1. Illustration of Huygen's principle applied to fire growth modelling under various fuel types and wind-slope vectors

Other physics-based simulations, such as Mell et al. [19], harness their associated weakness of data by hyper focusing on specific fire spreading environments. In this specific case, grassland fires were modelled favorably using known equations from fluid dynamics, combustion, and thermal degradation.

Finally, WRF-SFIRE [18] is a predictive weather forecasting model coupled with a fire-spread model (SFIRE). WRF-SFIRE is based on the Clark-Hall mesoscale atmospheric model with a tracer-based fire spread model. Basic evaluation revealed conclusive predictive ability with further test-time evaluations yet to be conducted.

Though these models may be effective in controlled settings, they commonly require unrealistically precise measurements from fuel distribution and moisture to elevation data.

2.2. Supervised Learning

The dataset itself presents a simple neural network with a convolutional encoder as a modelling technique [14], achieving a precision and recall of 33.6% and 43.1% respectively. A similar approach was taken by Shadrin et al. [28] who used CNNs in a U-Net style, specifically in the MA-Net style [13]. This approach achieved an exceptional F1 score of 0.85, though it was highly optimized for particular regions in the Republic of Sakha.

Finally, an alternative vision based method detailed in Wu et al. [30] uses a two layer neural network to generate a flame propagation map which subsequently uses a physical fire propagation model to predict fire spread. This model achieved equivalently exceptional F1 scores of 89.85%, similar to Shadrin et al. [28], but was similarly limited in scope to regions of China.

2.3. Limitations

An important observation on existing computer vision solutions predicting wildfire spread is the disparity between training and test data. The geographic and environmental variables used during training are often rich and high resolution even though these AI systems may encounter partially observed datapoints or high noise during deployment [33]. Robustness within these extreme circumstances are often the most critical and useful to first responders, prompting this paper to investigate prediction models under regimes with limited observability.

3. Dataset

Next Day Wildfire Spread [14] is a Kaggle based dataset containing 18,545 samples aggregated across the contiguous United States from 2012 to 2020. Each sample contains various explanatory variables describing a 64 km by 64 km region with 1 km resolution. Next Day Wildfire Spread's explanatory variables include, but are not limited to: topography, vegetation, weather, drought index, and population density. See Figure 2 for example datapoints.

Importantly, each sample contains a previous-day fire segmentation mask and a corresponding next-day gold label fire segmentation mask with a noted temporal separation of a day. These segmentation masks contain three distinct classes: -1 (no data), 0 (no fire), and 1 (fire). Class 0 (no fire) overwhelmingly dominates these segmentation masks in terms of pixel count with a pixel frequency of 97.26% contrasting the pixel frequency of no data (1.77%) and of fire-containing pixels (0.98%).

Since our model aims to predict fire spread irrespective of observability, we only predict on the two data classes (0/no fire and 1/fire). Data was split into 14,979 train examples, 1,689 test examples, and 1,877 validation examples.

4. Methods

4.1. Loss

Given the aforementioned class imbalance present in the fire segmentation masks (background pixel percentage of 97.26% for no-fire pixels compared to a foreground pixel percentage of 2.74% for data relevant classes), we require

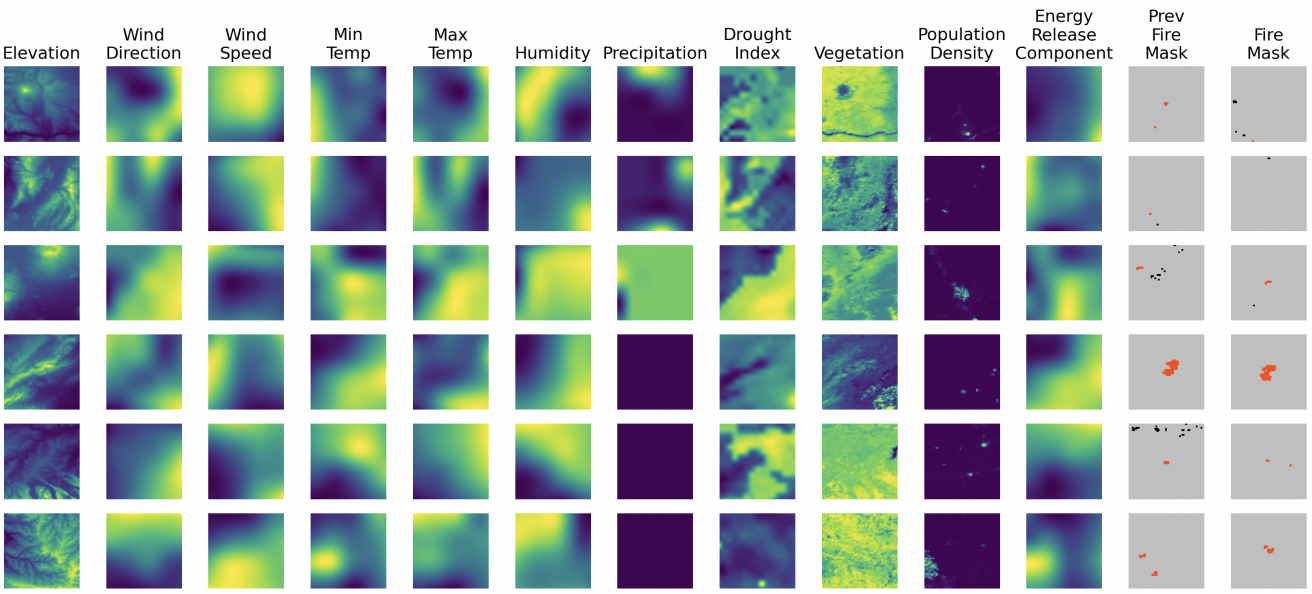

Figure 2. Six sampled datapoints taken from **Next Day Wildfire Spread dataset** showing associated features. Note that on the fire segmentation masks black pixels denote no data (-1), gray pixels denote no fire (0) and orange pixels denote fire (1).

a loss function that can weigh the classes according to their frequency to mitigate any artificially low losses. Our algorithm modifies the Dice score coefficient into a loss.

$$\text{Dice} = \frac{2\sum_i(\hat{y}_i y_i)}{\sum_i(\hat{y}_i)^2 + \sum_i(y_i)^2} \tag{1}$$

The Dice score coefficient (1) is a similarity metric that compares two binary arrays using overlap, innately accounting for class imbalance. It only compares the positive intersection, effectively ignoring all 0 classes. Here $\hat{y}_i$ is the predicted pixel value and y_i is the real pixel value. Notably, when applied to binary masks, the Dice coefficient is identical to the F1 score. This assumes pixel values only take on classes 0 and 1. Since we want to drive similarity up and since the coefficient is always less than 1, we can convert this to a loss in a simple way.

$$\begin{aligned} \text{Loss}_{\text{Dice}} &= 1 - \text{Dice} \\ &= 1 - \frac{2\sum_i(\hat{y}_i y_i)}{\sum_i(\hat{y}_i)^2 + \sum_i(y_i)^2} \end{aligned} \tag{2}$$

It is important to note that our model is predicting only two classes even though the gold label fire segmentation mask may contain three. In actuality, our model is masking a particular class when calculating loss: the -1 class (no data).

While pure Dice loss handles class imbalances and provides good signal for segmentation accuracy, it has been found to be susceptible to adversarial perturbations [24]. To increase robustness, Dice loss can be combined with binary cross entropy (BCE) (3) loss. To mitigate the effects of class imbalances, BCE loss can be modified to weighted binary cross entropy (WBCE) (4) which accounts for the distribution of classes by weighting the binary classes.

$$\begin{aligned} \text{BCE}_i &= -(y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)) \\ \text{Loss}_{\text{BCE}} &= \frac{1}{N}\sum_{i=0}^{N} \text{BCE}_i \end{aligned} \tag{3}$$

$$\text{Loss}_{\text{WBCE}} = \frac{1}{N}\sum_{i=0}^{N}(y_i \cdot w_1 + (1 - y_i) \cdot w_0)\text{BCE}_i \tag{4}$$

Where w_1 is the weight associated with class 1 (fire) and w_0 is the weight associated with class 0 (no fire). Our loss sets $w_1 = 100$ and $w_0 = 1$ matching the frequency distribution of pixel classes. These loss functions can be combined using a weighted sum, giving us our final loss function (5).

$$L = \text{WBCE} + 2 \cdot \text{Dice} \tag{5}$$

4.2. Evaluation

Our model leverages three evaluation metrics commonly used throughout the literature to determine predicted segmentation mask accuracy: mean intersection over union (Mean-IoU), mean pixel accuracy (MPA), and average distance [20]. Note that F1 score was not used since the loss function (Dice loss) is identical to F1 score when applied to a binary mask.

IoU is the area of the intersection between predicted and gold label segmentation mask over the union of these two segmentation tasks (6). Mean-IoU is the IoU averaged over

all classes. IoU is often considered better at penalizing under and over segmentation [21].

$$\begin{aligned} \mathrm{I}_c &= \sum_i \mathbb{1}_{y_i=\hat{y}_i=c} \\ \mathrm{IoU}_c &= \frac{\mathrm{I}_c}{\sum_i \mathbb{1}_{y_i=c} + \sum_i \mathbb{1}_{\hat{y}_i=c} - \mathrm{I}_c} \\ \text{Mean-IoU} &= \frac{1}{|C|} \sum_{c \in C} \mathrm{IoU}_c \end{aligned} \tag{6}$$

MPA is a trivial accuracy measurement involving the number of correctly classified pixels over the total number of associated pixels, averaged over every class (7).

$$\begin{aligned} \mathrm{PA}_c &= \frac{\sum_i \mathbb{1}_{y_i=\hat{y}_i=c}}{\sum_i \mathbb{1}_{y_i=c}} \\ \mathrm{MPA} &= \frac{1}{|C|} \sum_{c \in C} \mathrm{PA}_c \end{aligned} \tag{7}$$

Since the predicted fire segmentation mask is made up of probabilities, an insightful evaluation metric would be the proximity these probabilities are to correctly describing gold label segmentation masks. We can compute the average distance between probabilities and gold labels using a simple euclidean norm (8). This metric provides valuable insights into the models' ability to produce accurate predictions.

$$\text{Average Distance} = \sqrt{\sum_i (y_i - \hat{y}_i)^2} \tag{8}$$

4.3. Baseline

The baseline architecture took inspiration from the **Next Day Wildfire Spread** Dataset's baseline [14]. This baseline involved implementing a MobileNetV2 encoder layer followed by a seven layer decoder built up of six upsample blocks completed with a convolutional layer (4.3).

MobileNetV2 utilizes inverted residual architectures to allow the model to learn and embed interesting features into a higher dimension which subsequently gets collapsed back down once learning has been completed [26]. Further complexity is added during this projection with the addition of residual connections between bottleneck layers to improve gradient flow and training efficiency.

4.4. Feature Rich

Our feature rich architecture aims to exploit as much information as possible from the twelve explanatory variables (irrespective of the foreshadowed deviations we may experience during deployment) with performance in mind. This architecture took inspiration from the widely used Mask R-CNN segmentation model [11] which uses a regional CNN (R-CNN) with supplemented mask heads (4.4).

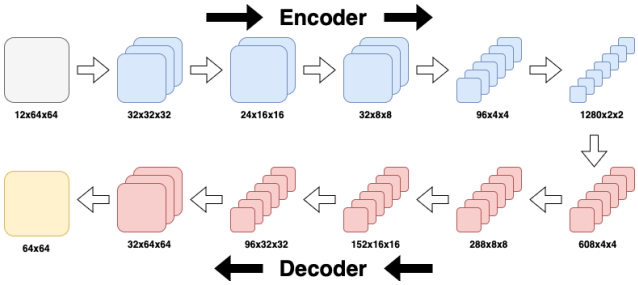

Figure 3. Baseline architecture inspired by the **Next Day Wildfire Spread** dataset. All activation functions are ReLU and the optimizer chosen is AdamW. MobileNetV2's first convolutional layer is modified to take in all twelve features and the final upsampling blocks are completed with a convolutional layer to rescale to the appropriate 64x64 image resolution.

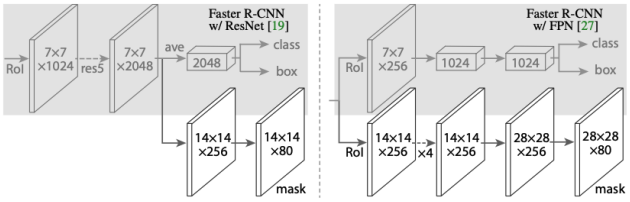

Figure 4. Head architecture for Mask R-CNN detailing the mask branches added to ResNet C4 and FPN backbones. The **Feature Rich** model built off of this backbone architecture.

Our architecture composed two known architectural patterns to create a Mask R-CNN adjacent segmentation model. The first pattern contained two 3x3 convolution layers designed to upscale the number of channels while keeping number of parameters constant. This design choice was meant to allow the model to decipher increasingly complex features to better predict the potential of fire spread. Downscaling was provided by two convolution layers, one 3x3 convolution layer and a 1x1 convolution layer used to effectively reduce the high channel dimension into a more manageable form to be fed into the ResNet50 backbone [12]. From there, a fully connected layer was used to reduce the dimension to the expected 64x64 resolution.

4.5. Data Restricted

In contrast to the **Feature Rich** architecture explained previously, the data restricted model aims to perform irrespective of the environmental conditions specifically by assuming as little data as possible. This is analogous to a worst-1 situation as described in the literature where virtually no data is accessible [33]. For this architecture, only three features were chosen for propagation: elevation, vegetation, and population density. We chose these features with accessibility in mind; in other words, we believe these features could be generated simply given a satellite image so that during deployment wildfires can be accurately predicted only with the use of a dedicated satellite.

With these restrictions in mind, this architecture reflects the **Feature Rich** model in regards to choice of upscaling and downscaling using convolutional layers. Our data restricted model uses two 3x3 convolutional layers to upscale with max pooling and dropout added during upscaling to improve regularization and numerical stability. For downsampling, again one 3x3 and one 1x3 convolution layer was used with max pooling and dropout used as well. This was finalized with a fully connected layer.

5. Results

Hyperparameters were explored during the initial phase of experimentation. It was determined via early accuracy measurements that the best hyperparameters for experimentation are learning rate of 10^{-3}, batch size of 32, and the use of the AdamW optimizer. Such a low batch size was used to increase model specificity with the other two hyperparameters being chosen for their documented positive effects on convergence. Generally around 15 to 50 epochs were used for training.

Generally, we see similar performance across all models with the data restricted model edging out both the baseline and the **Feature Rich** model on accurately classifying future fire spread. With that being said, the overall best performing model in terms of Mean-IoU and MPA is the **Feature Rich** model with the lowest average distance coming from the baseline.

5.1. Loss

The overall structure of these models' loss functions during training is not standard. We see plateauing loss with little to no decrease in loss over the training barring for stochasticity (5). Loss that plateaus is often caused by two factors: incorrect hyperparameters or a model that is ineffective (either due to over complexity or a lack of complexity). A range of hyperparameters, specifically learning rate and batch size changes, were explored during the initial phase of testing. These tests yielded the same results, plateauing loss, which ultimately prompting us to believe the cause of this error comes from model concerns. The most likely culprit is that the lack of computational resources (specifically time allowed for training) ultimately led to an apparently flat loss even though in actuality the loss is decreasing but the time scale makes is such that this decrease is is too small to see.

With this being said, we see that the validation loss changes over the training time scale. This is evidence that some effective (or ineffective in the case of the limited data regime) learning is being done. To further investigate this, a random set of test points were predicted upon and their associated fire segmentation masks were generated (6). These predictions glaringly display some form of reward hacking where every model has picked up on the generalized data

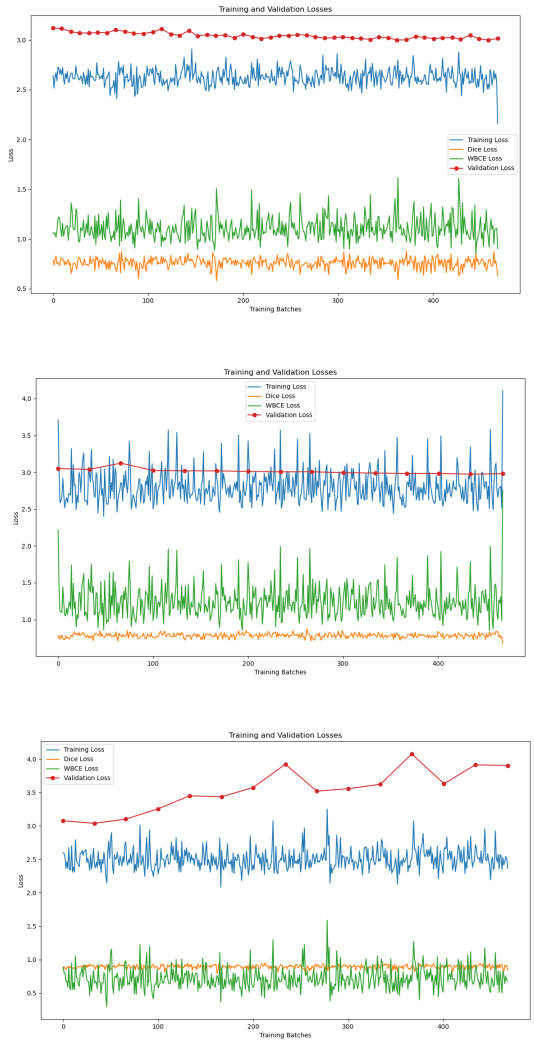

Figure 5. Loss plots over the training of the three models: baseline (top), feature rich (middle), and data limited (bottom). Notice the plateau of each loss barring stochasticity yet the either subtly increasing or decreasing validation loss.

bias for fire to most likely be present in the middle of the image. Consequently, every model predicts some general fire spread within the middle of the image irrespective of the inputted explanatory variables. This reward hacking is not strictly caused by a faulty loss function and may be caused simply by an unrepresentative dataset, though future work is required to explore whether there exists some metric to negate this form of model exploitation.

5.2. Confusion Matrices

From the faulty seeming loss functions presented above, it may seem evident that our models have failed to learn any useful functionality. Yet, we clearly see positive examples

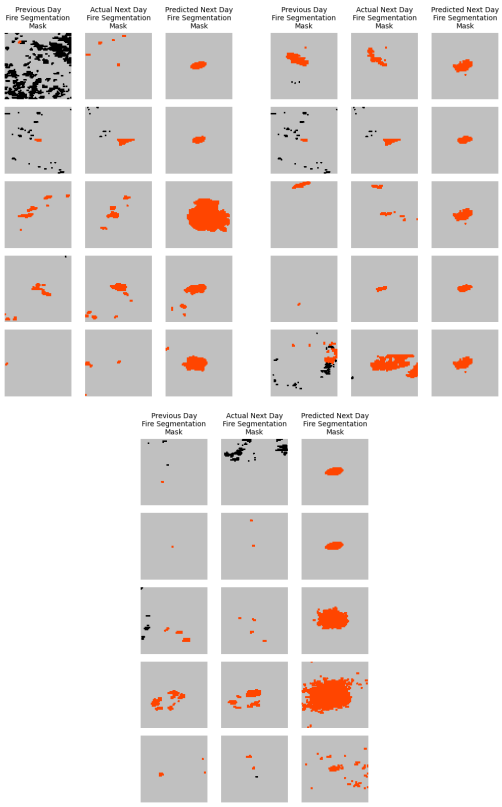

Figure 6. Fire segmentation mask predictions for the three models: baseline (top left), feature rich (top right), data limited (bottom). Notably, all three models appear to have picked up on bias present in the data for the common occurrence of fire spread in the center of each next-day fire segmentation mask. Note that in each example the left column is the previous day segmentation mask, the middle column is the true next-day segmentation mask, and the right column is the model-specific predicted next-day segmentation mask.

of prediction on unseen datapoints as shown in the fourth row of the baseline predictions and the fifth row of the data limited predictions in Figure 6. To further explore the ability of our models to predict fire spread, confusion matrices (7) were used to cross check their ability to positive predict class 1 examples. Across every model, there is an overwhelmingly positive ability to correctly classify class 0 as is expected given the class imbalance previously discussed between class 0 and class 1. Given the class imbalance considerations made in determining the loss function, these confusion matrices provide evidence that the model has not hacked this imbalance and is indeed attempting to predict segmentation masks.

Investigating the ratio between true positives and false negatives (7), we see that both the baseline and feature rich model are drastically underperforming with respect to their ability to positively classify a true class 1 pixel as class 1 or class 0. With the baseline model, the probability of it correctly classifying a true class 1 pixel as class 1 is 39.39% and with the feature rich model its probability is even worse at 30.58%. Given that there are only two classes, any probability less than 50% is considered abysmal. This is in comparison to the data limited regime, which predicts a true class 1 pixel as class 1 with a probability of 72.20%. While this seems promising, it is evident that the data limited regime over compensates and classifies nearly ten times as many pixels as positive pixels when compared to baseline and feature rich settings. Yet, considering how little data was provided to the data limited setting, this ability to accurately classify a positive pixel irrespective of the overestimation present in the model is impressive.

5.3. Accuracy Metrics

We see dominant performance by the feature rich model when using the accuracy metrics we defined for this task. The feature rich model peaks amongst all three models at both Mean-IoU and MPA, arguably the two most important heuristics of performance considering their ability to control for data imbalance and incorrect classification as well as positive classification (1).

Though the feature rich model achieves peak Mean-IoU and MPA on test data, we see an interesting pattern emerge when analyzing the output prediction masks from the feature rich model (6): repeated prediction patterns as seen in rows one, three, and five. This goes to show that perhaps the feature rich model is simply overfitting to training data and this performance may be artificial.

With that being said, the baseline model achieves a similarly high performance to the feature rich model across the board with a notably low average distance metric. Given how distinct both the baseline and feature rich models are from one another, these models are unlikely to achieve such similar accuracies, suggesting that more computational resources should be allocated to determine the effectiveness of these architectures.

As previously discussed, even though the data limited regime's accuracy metrics are underwhelming, other statistics and manual inspections suggest that this model has performed well given the lack of data availability provided to it.

	Mean-IoU	**MPA**	**Average Distance**
Baseline	0.55175	0.9683	**61.2018**
Feature Rich	**0.55367**	**0.9749**	180.8642
Data Limited	0.44404	0.8372	124.3158

Table 1. Accuracy metrics across all three models

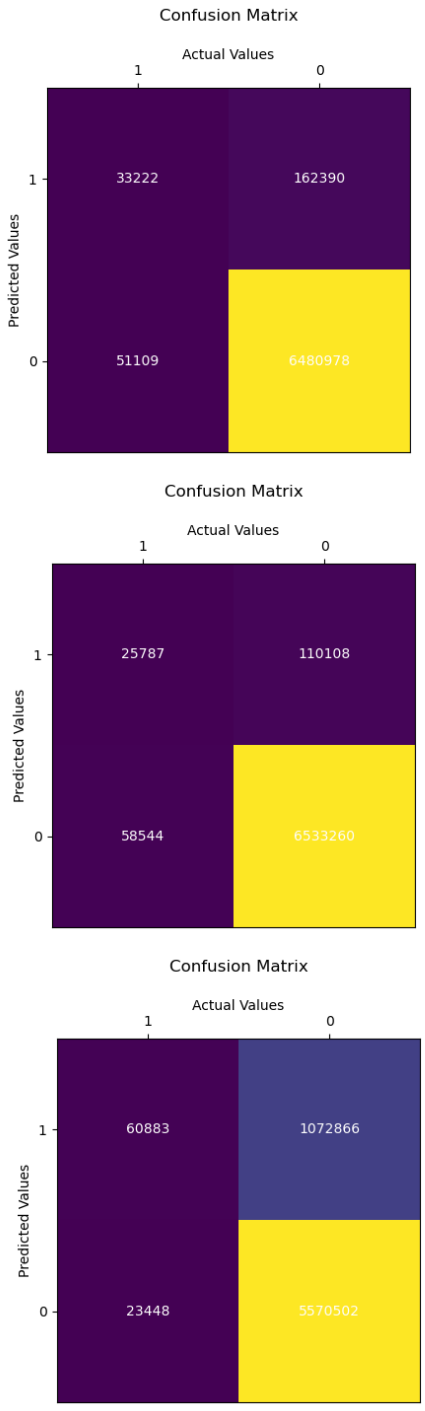

Figure 7. Confusion matrices over the training of the three models: baseline (top), feature rich (middle), and data limited (bottom). Notice the ease in each models' ability to predict class 0 labels yet the difficulty of the model to predict class 1 labels. In the data limited regime, a surprising majority of class 1 labels are accurately classified.

6. Conclusions/Future Work

Even though additional explanatory variables can increase accuracy during deployment, we have seen that these explanatory variables are not vital to the successful prediction of wildfire spread. Specifically, if we assume data was only collected from a satellite (as is done in the data restricted mode), relatively similar accuracy to the feature rich mode can still be gathered at the cost of overestimation.

Future work can be done to pair the explanatory variables used in the **Next Day Wildfire Spread** dataset with satellite imagery. Often detailed explanatory variables are unknown when calculating fire spread during deployment; adding satellite data to our model would allow for maximal robustness since this is the most easily accessed data. In addition to this dataset change, basic image manipulations to create synthetically new data was not explored due to already existing computational complexity concerns.

Potential architectural changes include more advanced techniques such as meta learning, ensemble of experts, or physics-backed vision solutions (such as [30]). Given the computational limitations of this project, these architectures have not been explored though are plausible avenues for future work given the distinct features presented and the overall task structure of predicting fire spread.

7. Contributions and Acknowledgements

Baseline CNN model and data retrieval code was inspired from the Kaggle **Next Day Wildfire Spread** dataset [14]. All other work was done manually and can be found at `github.com/jmichaels32/fireprediction`.

References

[1] J. T. Abatzoglou, A. P. Williams, and R. Barbero. Global emergence of anthropogenic climate change in fire weather indices. *Geophysical Research Letters*, 46(1):326–336, 2019. 1

[2] P. A. Arias, J. A. Rivera, A. A. Sörensson, M. Zachariah, C. Barnes, S. Philip, S. Kew, R. Vautard, G. Koren, I. Pinto, M. Vahlberg, R. Singh, E. Raju, S. Li, W. Yang, G. A. Vecchi, and F. E. L. Otto. Interplay between climate change and climate variability: the 2022 drought in central south america. *Climatic Change*, 177(1):6, Dec 2023. 1

[3] D. F. Balting, A. AghaKouchak, G. Lohmann, and M. Ionita. Northern hemisphere drought risk in a warming climate. *npj Climate and Atmospheric Science*, 4(1):61, Dec 2021. 1

[4] S. J. Brown. Future changes in heatwave severity, duration and frequency due to climate change for the most populous cities. *Weather and Climate Extremes*, 30:100278, 2020. 1

[5] M. Carreras-Sospedra, S. Zhu, M. MacKinnon, W. Lassman, J. D. Mirocha, M. Barbato, and D. Dabdub. Air quality and health impacts of the 2020 wildfires in california. *Fire Ecology*, 20(1):6, Jan 2024. 1

[6] M. H. Eldesouki, A. E. Rashed, and A. A. El-Moneim. A comprehensive overview of carbon dioxide, including emission sources, capture technologies, and the conversion into value-added products. *Clean Technologies and Environmental Policy*, 25(10):3131–3148, Dec 2023. 1

[7] M. A. Finney. *FARSITE: Fire Area Simulator-model development and evaluation*. 1998. 2

[8] M. Flannigan, A. S. Cantin, W. J. de Groot, M. Wotton, A. Newbery, and L. M. Gowman. Global wildland fire season severity in the 21st century. *Forest Ecology and Management*, 294:54–61, 2013. The Mega-fire reality. 1

[9] F. M. T. Force. Jan 2021. 1

[10] G. Forzieri, L. Feyen, S. Russo, M. Vousdoukas, L. Alfieri, S. Outten, M. Migliavacca, A. Bianchi, R. Rojas, and A. Cid. Multi-hazard assessment in europe under climate change. *Climatic Change*, 137(1):105–119, Jul 2016. 1

[11] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick. Mask R-CNN. *CoRR*, abs/1703.06870, 2017. 4

[12] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. 4

[13] K. Hettihewa, T. Kobchaisawat, N. Tanpowpong, and T. H. Chalidabhongse. Manet: a multi-attention network for automatic liver tumor segmentation in computed tomography (ct) imaging. *Scientific Reports*, 13(1):20098, Nov 2023. 2

[14] F. Huot, R. L. Hu, N. Goyal, T. Sankar, M. Ihme, and Y. Chen. Next day wildfire spread: A machine learning data set to predict wildfire spreading from remote-sensing data. *CoRR*, abs/2112.02447, 2021. 1, 2, 4, 7

[15] M. Jerrett, A. S. Jina, and M. E. Marlier. Up in smoke: California's greenhouse gas reductions could be wiped out by 2020 wildfires. *Environmental Pollution*, 310:119888, 2022. 1

[16] M. Kabir, U. E. Habiba, W. Khan, A. Shah, S. Rahim, P. R. D. los Rios-Escalante, Z.-U.-R. Farooqi, L. Ali, and M. Shafiq. Climate change due to increasing concentration of carbon dioxide and its impacts on environment in 21st century; a mini review. *Journal of King Saud University - Science*, 35(5):102693, 2023. 1

[17] C. D. Keeling. Climate change and carbon dioxide: an introduction. *Proc Natl Acad Sci U S A*, 94(16):8273–8274, Aug. 1997. 1

[18] J. Mandel, J. D. Beezley, and A. K. Kochanski. Coupled atmosphere-wildland fire modeling with wrf 3.3 and sfire 2011. *Geoscientific Model Development*, 4(3):591–610, 2011. 2

[19] W. Mell, M. A. Jenkins, J. Gould, and P. Cheney. A physics-based approach to modelling grassland fires. *International Journal of Wildland Fire*, 16(1):1–22, 2007. 2

[20] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos. Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3523–3542, 2022. 3

[21] D. Müller, I. Soto-Rey, and F. Kramer. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res Notes*, 15(1):210, June 2022. 4

[22] C. Nelson. Assessing the efficacy of california's wildfire and forest resilience action plan. 2022. 1

[23] A. D. of Natural Resources. March 2004. 1

[24] V. Rajput. Robustness of different loss functions and their impact on networks learning capability, 2021. 3

[25] B. H. Samset, C. Zhou, J. S. Fuglestvedt, M. T. Lund, J. Marotzke, and M. D. Zelinka. Steady global surface warming from 1973 to 2022 but increased warming rate after 1990. *Communications Earth & Environment*, 4(1):400, Nov 2023. 1

[26] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L. Chen. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. *CoRR*, abs/1801.04381, 2018. 4

[27] F. Sari. Identifying anthropogenic and natural causes of wildfires by maximum entropy method-based ignition susceptibility distribution models. *Journal of Forestry Research*, 34(2):355–371, Apr 2023. 1

[28] D. Shadrin, S. Illarionova, F. Gubanov, K. Evteeva, M. Mironenko, I. Levchunets, R. Belousov, and E. Burnaev. Wildfire spreading prediction using multimodal data and deep neural network approach. *Scientific Reports*, 14(1):2606, Jan 2024. 2

[29] L. Wang, L. Wang, Y. Li, and J. Wang. A century-long analysis of global warming and earth temperature using a random walk with drift approach. *Decision Analytics Journal*, 7:100237, 2023. 1

[30] Z. Wu, B. Wang, M. Li, Y. Tian, Y. Quan, and J. Liu. Simulation of forest fire spread based on artificial intelligence. *Ecological Indicators*, 136:108653, 2022. 2, 7

[31] R. Xu, P. Yu, Y. Liu, G. Chen, Z. Yang, Y. Zhang, Y. Wu, P. J. Beggs, Y. Zhang, J. Boocock, F. Ji, I. Hanigan, O. Jay, P. Bi, N. Vargas, K. Leder, D. Green, K. Quail, R. Huxley, B. Jalaludin, W. Hu, M. Dennekamp, S. Vardoulakis, A. Bone, J. Abrahams, F. H. Johnston, R. Broome, T. Capon, S. Li, and Y. Guo. Climate change, environmental extremes, and human health in australia: challenges, adaptation strategies, and policy gaps. *Lancet Reg Health West Pac*, 40:100936, Nov. 2023. 1

[32] A. M. Young, P. E. Higuera, P. A. Duffy, and F. S. Hu. Climatic thresholds shape northern high-latitude fire regimes and imply vulnerability to future climate change. *Ecography*, 40(5):606–617, 2017. 1

[33] J. Zhou, W. Jiang, F. Wang, Y. Qiao, and Q. Meng. Comparing accuracy of wildfire spread prediction models under different data deficiency conditions. *Fire*, 7(4), 2024. 1, 2, 4