Predicting Forest Fire Using
Remote Sensing Data And Machine Learning

Suwei Yang, [1] Massimo Lupascu, [2] Kuldeep S. Meel [1]

[1] School of Computing, National University of Singapore
[2] Department of Geography, National University of Singapore

Abstract

Over the last few decades, deforestation and climate change have caused increasing number of forest fires. In Southeast Asia, Indonesia has been the most affected country by tropical peatland forest fires. These fires have a significant impact on the climate resulting in extensive health, social and economic issues. Existing forest fire prediction systems, such as the Canadian Forest Fire Danger Rating System, are based on handcrafted features and require installation and maintenance of expensive instruments on the ground, which can be a challenge for developing countries such as Indonesia. We propose a novel, cost-effective, machine-learning based approach that uses remote sensing data to predict forest fires in Indonesia. Our prediction model achieves more than 0.81 area under the receiver operator characteristic (ROC) curve, performing significantly better than the baseline approach which never exceeds 0.70 area under ROC curve on the same tasks. Our model's performance remained above 0.81 area under ROC curve even when evaluated with reduced data. The results support our claim that machine-learning based approaches can lead to reliable and cost-effective forest fire prediction systems.

1 Introduction

Over the last few decades, forest fires, especially peatland forest fires, have been a major environmental issue in Southeast Asia, with significant impacts on the atmosphere, carbon cycle and various ecosystem services (Page et al. 2009). Forest fires can affect a large portion of the population, causing economic difficulties and disruption to businesses and short and long term health problems (Jaafar and Loh 2014; Chisholm, Wijedasa, and Swinfield 2016; Huijnen et al. 2016; Stockwell et al. 2016). The smoke and haze produced by the fires can also cause disruptions and economic losses to neighbouring countries in hundreds of millions (Doerr and Santín 2016; World Bank 2016). Thus it is important to be able to predict forest fires and take mitigating measures to reduce their devastating impacts.

We focus on Indonesia because it accounts for 83% of total peatland forests by area in Southeast Asia (Page and Rieley 2016). Peatland forests are important primarily because of their function as carbon storage (Page, Rieley, and

Banks 2011). If peatland forests are burnt, the stored carbon would be released into the atmosphere as CO_2, CH_4, CO, etc. Indonesia is estimated to have 28.1 gigatons of carbon pool (Warren et al. 2017) in its peatlands. Whenever there is a large scale forest fire, the effects are devastating as shown by previous events of 1998, 2015 and 2019. In 2015, 4.6M hectares were burned and 0.89 gigatons of carbon dioxide equivalent were released (Lohberger et al. 2018).

Existing forest fire prediction systems mainly take a modeling approach and use handcrafted features to perform predictions (Stocks et al. 1989; MA 2004; San-Miguel-Ayanz et al. 2012). One such existing system is the Canadian Forest Fire Danger Rating System (Stocks et al. 1989), which is designed to evaluate environmental factors that influence the ignition, spread, and behavior of wildland fire. The system is managed by the Canadian government and has been adopted by other organizations around the world such as the New Zealand Forest Service (Fogarty et al. 1998) and European Commission which oversees the European Forest Fire Information System (EFFIS) (López, San-Miguel-Ayanz, and Burgan 2002).

Modeling-based approaches with handcrafted features and closed-form equations suffer from two drawbacks. Firstly, comprehensive fire prediction models have to deal with numerous weather-related and soil variables and their complex inter-relationships. It is difficult to describe all these intricate relationships with a set of mathematical equations. Consequently, scientists working on fire prediction systems approximate relationships between various weather-related factors by introducing simplifying assumptions (Kloprogge, van der Sluijs, and Petersen 2011). This might lead to a lack of robustness in the predictions, requiring system recalibration according to different deployment environments. Secondly, the models tend to rely on data from instruments on the ground and this leads to high operating costs. For example, the Canadian Forest Fire Danger Rating System operates on data from over 750 weather stations on the ground. Operating a similar system may be challenging in terms of cost for a developing country such as Indonesia. In light of the previously mentioned drawbacks coupled with recent advances in machine learning techniques and availability of remote sensing data, we want to answer *Can machine-learning based techniques be used for forest fire prediction using remote sensing data?*

In this work, we use supervised learning to train a neural network with remote sensing data to predict forest fires in Indonesia. In particular, the remote sensing data consists of historical Landsat 7 satellite images (U.S. Geological Survey 2019) and Fire Information for Resource Management System (FIRMS) (Davies et al. 2009) hotspot dataset. We designed our model to predict forest fires one month into the future, in line with the prediction time-frame of prior work (Maulana et al. 2019). The time-frame requirement for prediction into the future stems from the need for time to plan resource intensive operations in remote terrains to either respond to potential forest fires (Maulana et al. 2019) or prevent fires in the first place. More formally, at time t, our model takes in a year of historical data (i.e., sliding window of data belonging to $[t - 52\text{week}, t]$) and outputs confidence of the presence of fire for $t + 5\text{week}$, which is referred to as *fifth week into future* for rest of the paper.

We performed extensive empirical evaluations on the prototype implementation of our system, called Agni[1]. In particular, Agni achieves more than 0.81 area under the receiver operator characteristic (ROC) curve in our evaluations. We also performed comparative studies with the logistic regression-based baseline that was proposed in prior work (Maulana et al. 2019). In our evaluations, Agni performed significantly better than the baseline. The baseline approach never exceeded 0.70 area under ROC curve while Agni always stayed above 0.80.

Given the need for cost-effectiveness, we further evaluated Agni with reduced data. We limited the duration of historical data available to Agni for training and prediction to just 3 months and its performance still remained above 0.81 area under the ROC curve. In summary, our empirical evaluation demonstrates that Agni, a machine-learning based approach using remote sensing data, is effective while being more cost-effective than existing systems that required instruments on the ground.

The rest of the paper is structured as follows. First, we discuss related prior works in Section 2, then we provide details of the data used and its challenges in Section 3. In Section 4, we discuss the details of our approach and the challenges encountered. Subsequently, we provide details of the evaluation in Section 5 and conclude in Section 6.

2   Related Work

Existing forest fire forecasting systems, such as the popular Canadian Forest Fire Weather Index System (Wagner 1987), rely on data from weather stations (over 750 in this case) on the ground to operate. For the Canadian Forest Fire Weather Index System, the input data is then processed according to closed-form mathematical equations to attain Fire Weather Index (FWI) which is a numerical rating of fire intensity. As discussed above, our work seeks to develop alternative techniques based on publicly available satellite image data to provide cost-effective prediction systems for developing countries such as Indonesia.

While our work is the first in its kind to use neural network with remote sensing data to predict forest fires, a related

---

[1] code available at https://github.com/meelgroup/agni

study by Maulana et al. (2019) tried to predict peatland forest fires using a spatial logistic regression approach. Specifically, Maulana et al. employ careful expert-driven manual feature engineering involving features such as river network density and peat soil depth. Such features, while useful in their smaller case study, are not readily available across Indonesia. In contrast, our model takes only the satellite images as input, and is not restricted to peatland forest fires but predicts for forest fires in general.

Apart from the study by Maulana et al., there are other existing machine-learning based approaches in the domain of forest fires but they do not specifically focus on fire prediction. Wijayanto et al. used Adaptive Neuro Fuzzy Interface System to predict whether a fire alarm trigger is a false positive and explored the relationship of fire with human activities (Wijayanto et al. 2017). Cortez and Morais used logistic regression, decision tree and support vector machines to predict the potential burn area given weather station data (Cortez and Morais 2007). FireNet performs real-time detection and annotation of fire boundaries in drone video feeds (Doshi et al. 2019). FireCast, on the other hand, tries to predict how a fire will spread in the near future (Radke, Hessler, and Ellsworth 2019).

3   Data

Supervised learning requires both input data and ground truth labels to train a model. We used Landsat 7 satellite images, retrieved from Google Earth Engine (Gorelick et al. 2017), as inputs to the neural network. For ground truth labels, we used FIRMS hotspot dataset and we refer to forest fires as hotspots for consistency with FIRMS. In this section, we discuss details of the data used and their associated challenges. Finally, we describe the data preparation process.

3.1   Landsat 7 Satellite Images

Band	Resolution	Channel	Wavelength (μm)
B1	30m	Blue	0.45-0.52
B2	30m	Green	0.52-0.60
B3	30m	Red	0.63-0.69
B4	30m	Near IR	0.77-0.90
B5	30m	IR 1 Shortwave	1.55-1.75
B6_1	60m	Low-gain IR 1	10.40-12.50
B6_2	60m	High-gain IR 1	10.40-12.50
B7	30m	Shortwave IR 2	2.09-2.35
B8	15m	Panchromatic	0.52-0.90

Table 1: Specifications of different bands on Landsat 7

Landsat 7 (U.S. Geological Survey 2019) satellite is part of the Landsat program, under NASA and U.S. Geological Survey. Landsat 7 has an orbit period of 16 days. Satellite images from Landsat 7 are used as the input source for our hotspot prediction model. The imaging instrument on Landsat 7 satellite, Enhanced Thematic Mapper Plus(ETM+), produces imagery of Earth as eight spectral bands. The satellite image specifications are as shown in Table 1. The Band column shows the respective names of the Landsat 7 satellite

bands. The Resolution column shows the resolution of the respective band. A resolution of 30m means that one pixel in the image band represents an area of $30m \times 30m$. The Channel column shows what each respective band represents and the Wavelength column shows the respective wavelengths of each band in μm.

Six of the bands have a resolution of 30m, except bands B6 and B8. However, band B6 is already resampled to the resolution of 30m in the Landsat 7 data available on Google Earth Engine. Since all bands except band B8 have the same resolution ($30m \times 30m$), we focus on usage of bands B1 to B7 in this work.

3.2 Fire Information for Resource Management System (FIRMS) Hotspot Data

In order to train a neural network for forest fire prediction, we need to provide ground truth labels of fire. We found the FIRMS hotspot dataset (Davies et al. 2009) to be the most comprehensive open fire dataset available. The FIRMS dataset, provided by NASA, is used by several existing fire prediction systems such as the Canadian Wildland Fire Information System, the European Forest Fire Information System and the U.S. Forest Service - Missoula Fire Sciences Lab. In this work, we use the FIRMS hotspot data as labels for training and evaluation of prediction models.

The FIRMS dataset consists of hotspots detected by LANCE [2] system after processing the MODIS [3] and VIIRS [4] instrument data on Terra, Aqua and Suomi-NPP satellites (National Aeronautics and Space Administration (NASA) 2021a,b,c,d,e; Karl Hille 2015). The resolution of the FIRMS dataset is 1km $\times$ 1km per hotspot pixel. Each detected hotspot instance is associated with a confidence level ranging from 0% to 100%, indicating certainty of the detection. For a given area, we assign the label to be the maximum confidence value within that area.
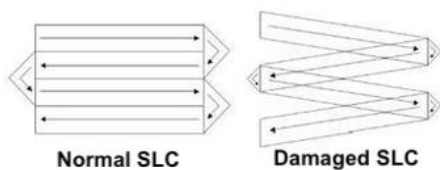
3.3 Challenges

Figure 1: Effect of damaged SLC (source: U.S. Geological Survey)

Imperfect Orbit and Damaged SLC We encountered two primary sources of missing data: imperfect orbit of satellite and damaged scan-line calibrator (SLC). The first

source of missing data arises from Landsat 7's imperfect orbit. Ideally, a satellite with a perfect orbit can produce images of the exact same area during each orbit. In reality, the Landsat 7 satellite images do not overlap perfectly with each other between orbits and sometimes parts of an area are missing due to imperfect orbit. To overcome the issue of imperfect image overlap, we choose a reference image and crop the remaining satellite images to the area of the reference image. When the satellite images are sliced up into smaller images during our data preparation process discussed later, some slices may be missing, i.e. entire image slice is filled with zero entries.

The other source of missing data mentioned earlier is Landsat 7's scan-line calibrator (SLC). The scan-line calibrator (SLC) ensures that the imaging instrument on Landsat 7 satellite scans in the correct manner when producing satellite images in a row-wise manner. A normal imaging motion with a working SLC is shown by the arrows on the left of Figure 1. When the SLC is damaged in the case of Landsat 7, the imaging motion is in a zigzag form, shown by the arrows on the right of Figure 1. The resulting satellite image produced by the damaged SLC has triangular strips of missing data.

For both sources of missing data encountered, we did not perform statistical imputation.[5] Our hypothesis is that the model should be able to learn from the remaining data. Our results show that the model is still able to perform well, supporting our hypothesis of the models learning ability on the remaining data. More details will be discussed in Section 5.

Resolution and Data Size Trade-off Ideally, we would want to use satellite image data at full resolution (30m by 30m per pixel for Landsat 7) for model training and evaluation. However, the capacity to handle full resolution images requires extensive storage. At full resolution, a reference satellite image area and its associated historical satellite images require around 2 gigabytes of storage. We used 800 such satellite image time series, requiring more than 1.6 terabytes of storage if full resolution satellite images are used. File size is still an issue, even when the image is downscaled. As a result, each image in a historical satellite image time series is first downscaled, then sliced into smaller images and subsequently converted to a histogram. In particular, each band of the image is converted to a histogram of 32 evenly spaced intervals that are referred to as bins, with each bin representing counts of 8 pixel values as each band has pixel value range from 0 to 255. For an image band b, the first bin in the histogram would be the total count of pixel values in band b, in the range of 0 to 7 inclusive. Since we work with 8 bands (B1 to B7), each image is converted into 8 histograms of 32 bins. The resultant histograms are a storage efficient representation of an 8km by 8km area.

3.4 Data Preparation

In the data preparation stage, retrieved historical satellite images are sliced into smaller images that represent 8km by
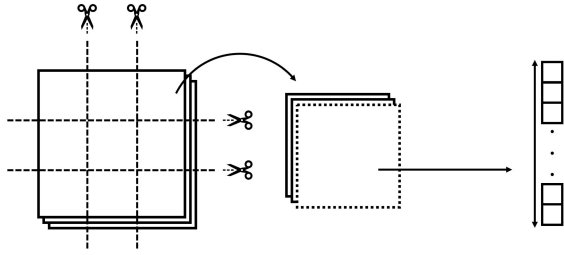
Figure 2: Data preparation process: The big satellite image time series is sliced into 8 x 8km area time series, then each channel is converted to histogram of 32 equal bins

---

Algorithm 1 dataPrep algorithm: algorithm to retrieve and preprocess satellite images and FIRMS labels. wk stands for week.

Input: t_s and t_e which represents start and end date
Output: Historical satellite image histogram and future labels (5th week into future)

1: refImgList $\longleftarrow$ getRefImgs(t_s, t_e)
2: **for** r_i in refImgList **do**
3:    t_{r_i} $\longleftarrow$ getDate(r_i)
4:    pastImgs $\longleftarrow$ getHistoricalImgs(r_i, $52wk$)
5:    label $\longleftarrow$ getFirmsLabel(r_i, $t_{r_i} + 4wk$, $t_{r_i} + 5wk$)
6:    imgLabelTensor $\longleftarrow$ combine(pastImgs, label)
7:    slicedDataList $\longleftarrow$ slice(imgLabelTensor, $8km$)
8:    **for** slicedData in slicedDataList **do**
9:      **if** !allZero(slicedData) **then**
10:       finalData $\longleftarrow$ toHist(slicedData)
11:       save(finalData)

8km areas. Subsequently, the sliced images are then converted to histograms and paired with the respective ground truth hotspot label from the FIRMS data. The data preparation process is illustrated in Figure 2 and Algorithm 1 provides further details about the process.

In line 1, the algorithm retrieves all reference satellite images of Indonesia within the interval specified by the input t_s and t_e. For each reference image, the past one-year of satellite images of the same area are retrieved in subroutine *getHistoricalImgs* on line 4. The *getHistoricalImgs* subroutine crops historical images to the exact area of the reference image. The *getFirmsLabel* subroutine on line 5 retrieves FIRMS hotspot data for the fifth week in the future with respect to the reference image date. The highest FIRMS value per pixel within the week long hotspot data is taken as the ground truth label. Line 6 combines the label with corresponding historical images. The combined image and label data are then sliced into smaller pieces to achieve the correct area of 8km by 8km in line 7. If an image is not entirely filled with zeros due to missing data (line 9), the subroutine *toHist* on line 10 converts each band of the sliced data into a histogram of 32 bins.

4 Approach and Challenges

We employed neural networks for the supervised learning task. We first present the architecture of the neural network employed and then shed light on two key design choices to handle the challenges specific to our application domain.

4.1 Neural Network Architecture

| Layer	Operation	Hidden Units
1	Conv2D (3×3)	64
2	LSTM	64
3	Dropout	-
4	Dense (ReLU)	256
5	Dense (ReLU)	32
Output	Dense (Sigmoid)	1

Table 2: Neural network architecture

The neural network architecture is shown in Table 2. The Layer column indicates the positions of layers in the architecture. The Operation column indicates the type of layer and the Hidden Units column indicates the number of hidden units in the respective layer. The neural network takes as input a histogram time series that represents an 8km by 8km area. The input is a one year time series of satellite image histograms with T timesteps. Each timestep consists of information from 8 bands, with 32 equally divided histogram bins for each band. The network first performs convolution (Fukushima 1980) on each input timestep of the data, which is the histogram of individual satellite images in the one year of historical satellite images. The convolution with 64 kernels of size 3×3 is applied to histogram bins of the different bands to learn representations of relationships between the different bands. The representation attained from the convolution layer, during each timestep, is passed to the Long-Short Term Memory (LSTM) layer (Hochreiter and Schmidhuber 1997) with 64 hidden units. The LSTM layer learns the temporal relationship of the historical data. This is followed by dropout with a rate of 0.3 and three dense layers with 256, 32 and 1 hidden units respectively. The last dense layer uses *sigmoid* activation to get an output value between 0 and 1. The output is interpreted as the predicted confidence of hotspot in that area for the fifth week into the future, from the most recent historical data timestep.
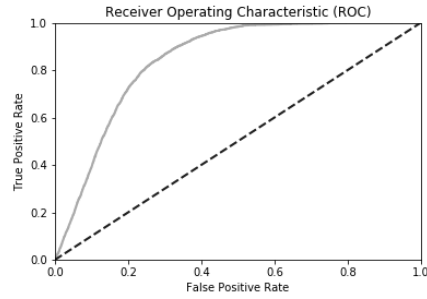
The neural network has a custom loss function L as follows:

$$L = |y_{true} - y_{pred}| \times$$
$$min(10^5, max(1, 10^{((y_{true} - y_{pred}) \times \frac{100}{k})}))$$

y_{true} is the label and y_{pred} is the predicted value for confidence of hotspot presence. Both y_{true} and y_{pred} take values between 0 and 1. In L, k is a hyper-parameter set to be 30 for best performance after hyper-parameter tuning. Other values of k that we have tried include 100, 50 and 10.
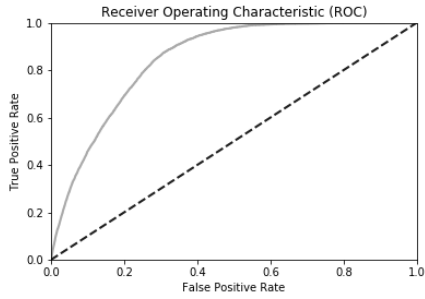
(a) September 2019 hotspots, AUC = 0.8480   (b) August 2019 hotspots, AUC = 0.8428

Figure 3: Area under ROC curve(AUC) for evaluation on (a)September and (b)August 2019 hotspots

The choice of loss function is primarily influenced by two factors: (1) the presence of skewed data, and (2) the cost of mis-predictions. As we will discuss in Section 4.2, the input data is heavily skewed towards absence of hotspots (i.e., y_{true} being mostly 0). The presence of skewed data requires us to steer the model away from always predicting y_{pred} to be close to 0, and therefore, we provide high penalty for the case of $y_{true} > y_{pred}$. Secondly, in practice, higher values of y_{pred} may give rise to false alarms but lower values of y_{pred} than the actual ground truth leads to unpreparedness and as such, leads to substantial loss of human, social, and economic capital.

4.2 Challenges

Heavily Skewed Data As briefly mentioned above, the natural distribution of the data is heavily skewed towards absence of hotspots. Empirical analysis of the data indicates that for a typical month, over 99% of locations do not have presence of hotspots. As such, a machine learning model may achieve 99% accuracy by simply predicting y_{pred} to be 0. In addition to the design of the custom loss function L discussed above, we perform the standard practice of re-sampling of the training data to achieve a 3:1 ratio for zero to non-zero labels.

Batch Size Limitations During Training As discussed in Section 3.3, the imperfect orbit of Landsat 7 leads to missing data for several timesteps when the data is viewed as a time series. The lack of data for certain timesteps has implications on the batch size for the training procedure of the neural network. In order to perform batch training, our data has to be of the same dimensions. However, that is not the case for our data because of missing timesteps in our data. Therefore, we train the neural network with a batch size of 1. An alternate approach would have been the usage of padding to ensure that the input training data instances within the same batch have the same dimensions. The lack of prior research about possible padding approaches for satellite image time series limited its use in our work but we believe exploring its role in future work is an interesting technical challenge.

5 Evaluation Setup and Results

To empirically validate our models, we implemented the prototype, Agni, using Tensorflow 1.12 and Python 2.7. We conducted the training and evaluation on high performance machines on compute clusters with the following specifications: Intel Xeon Silver 4108 with Nvidia Titan V and Intel Xeon Silver 4108 with Nvidia Tesla V100.

We employ a logistic regression approach as baseline, similar to prior work by Maulana et al.. In our baseline, we used mean squared error loss instead of binary cross entropy loss as the labels are real numbers from 0 to 1.

The choice of training dataset is influenced by three key considerations: (1) the need to minimize the amount of training data for computational efficiency, (2) the usage of data with less heavy skew towards the absence of hotspots, and (3) the usage of recent data so as to allow testing and validation of the model for recent months. Therefore, we employed data from 2014-18 as they coincided with large scale forest fires in Indonesia. We employed hotspots from April'18 as test set and hotspots from June'19 to September'19 as our validation set. Recall that we train Agni to predict labels based on historical data for the past year. In particular, the model is trained to predict labels for the 5th week ($t + 5$week) given a sliding window of one year of data, i.e., from $[t - 52$week$, t]$. This prediction period of 5th week into future is also in line with the spatial logistic regression work (Maulana et al. 2019). As mentioned in Section 4.2 the training data is of 3:1 ratio for zero to non-zero labels. Apart from 3:1 ratio, we also experimented with other ratios, specifically 4:1 and 2:1, but did not see a significant change in Agni's performance.

The objective of our evaluations was to answer the following questions:

RQ 1 How well does Agni predict hotspots?

RQ 2 How does the length of sliding window affect Agni's performance?

RQ 3 How does our custom loss function affect model performance?

5.1 Performance of Model (RQ 1)

For unambiguous communication with general public and across different agencies, we are often interested in the bi-

	Window			
Hotspot label	~3 Month	~6 Month	~9 Month	1 Year
June'19	0.8344	**0.8428**	0.8277	0.8125
July'19	0.8623	**0.8707**	0.8596	0.8519
Aug'19	0.8269	0.8471	**0.8511**	0.8428
Sept'19	0.8239	0.8464	0.8449	**0.8480**

Table 3: Area under ROC curve values for varying size of sliding window

	Model			
Hotspot label	Agni	Agni(MSE)	LR(custom loss)	LR(MSE)
June'19	**0.8125**	0.7483	0.6966	0.6355
July'19	**0.8519**	0.7081	0.7312	0.6967
Aug'19	**0.8428**	0.7264	0.7171	0.6613
Sept'19	**0.8480**	0.6956	0.7304	0.6560

Table 4: LR stands for logistic regression baseline and MSE stands for mean squared error loss function

narized output: whether there is going to be a hotspot or not. The standard process in machine learning community is to choose a threshold τ and binarize the output value based on τ. Any prediction value greater than τ is classified as *hotspot*, otherwise the prediction is classified as *no hotspot*. We interpret a non-zero label from FIRMS as *hotspot* and zero label from FIRMS as *no hotspot*. We plot the Receiver Operating Characteristic (ROC) curve, Figure 3, to show the behavior of our model across different τ values and we use area under ROC curve (AUC) as an indication of the model's performance. Figure 3 shows the ROC curves for August and September'19 and the AUC values for both August and September'19 are above 0.84. In this context, it is worth noting that if a model were to always predict no fire, the expected AUC would be 0.5, illustrated by the dotted lines in Figures 3a and 3b.

Hotspot label	Total labels	Positive hotspot labels
June'19	122416	495 (0.4%)
July'19	95594	1017 (1.1%)
Aug'19	104725	2885 (2.8%)
Sept'19	115934	6216 (5.4%)

Table 5: Distribution of labels of evaluation data. Postive hotspot labels - labels indicating non-zero confidence of hotspot presence. Total labels - total number of labels.

Further evaluations of the prediction model are conducted with June and July hotspots in 2019. The model had an AUC of 0.8519 and 0.8125 for prediction of July'19 and June'19 hotspots respectively. As evident from Table 5, there were significantly fewer hotspots for June and July'19 in comparison to August and September'19. The consistent performance of Agni across different months with varying distribution of data provides strong evidence in favor of robustness and versatility of the model. It is also worth noting that

our model performs significantly better than the logistic regression baseline as shown in Table 6. In the evaluations, Agni's AUC remained above 0.80 whereas the baseline's AUC never exceeded 0.70.

Hotspot label	Agni (our model)	LR baseline
June'19	**0.8125**	0.6355
July'19	**0.8519**	0.6967
Aug'19	**0.8428**	0.6613
Sept'19	**0.8480**	0.6560

Table 6: Area under ROC curve for Agni and LR baseline(logistic regression)

5.2 Impact of Sliding Window on Performance (RQ 2)

To understand the data requirements of Agni, we train and evaluate Agni with the length of sliding window varying from 3 months to 1 year. Table 3 shows the AUC values for the different lengths of sliding window used for training and prediction. Each column in Table 3 indicates the length of sliding window and each row corresponds to the month of hotspots predicted. For example, the cell corresponding to Sept'19 and ~3 Month shows the AUC value for Agni trained using a sliding window of the 6 most recent satellite images available and evaluated for Sept'19 hotspots.

5.3 Impact of Custom Loss Function on Performance (RQ 3)

To understand how much our custom loss function defined in Section 4.1 contributes to model prediction performance for $t + 5week$, we conducted ablation studies for the choice of loss function. We use the same evaluation data as before and compare against the performance of models in Table 6.

Specifically, we conducted additional evaluations on a verison of Agni with mean squared error as loss function and separately for a version of logistic regression model with our custom loss function. Our results, shown in Table 4, show that our custom loss function greatly improves the performance of both models and establishes its necessity. When we trained logistic regression using our custom loss function, its performance could exceed 0.70 AUC unlike before. Similarly, when we trained Agni using mean squared error, its performance dropped to lower than 0.75 AUC and in some cases worse than logistic regression trained with our custom loss function. This strongly supports our choice of custom loss function.

5.4 Summary

In summary, Agni trained with our custom loss function performs consistently better than the baseline model across different evaluation scenarios, as demonstrated by Tables 3 and 6, thereby providing strong evidence in support of the predictive power of machine-learning techniques using remote sensing data. Furthermore, Agni can be trained and evaluated with a significantly smaller sliding window, thereby requiring less storage of data during deployment and thus positioning itself as a cost-effective fire prediction system for developing countries such as Indonesia.

6 Conclusion

In this work, we trained a neural network to perform forest fire prediction. While existing prediction models are effective, they use data collected from instruments on the ground, leading to high operating costs. Our model uses readily available remote sensing data in the form of satellite images to predict fire hotspots. We showed that our model performs consistently well under extensive evaluations, providing an affirmative answer to the initial question *Can machine learning-based techniques be used for forest fire prediction using remote sensing data?*

An interesting area to explore in the future is using multiple data sources, such as precipitation, soil moisture and human activity data, to achieve better performance. In particular, human activity data may be useful for predicting fires started by humans. It is a common practice to use fire to clear land before planting crops, but the fire could spread to forest areas, becoming forest fires. The challenge in using multiple data sources would be integrating sources with different resolutions. This would require a redesigning of the neural network architecture to include multiple inputs and merging them within the network.

Acknowledgments

References

Chisholm, R. A.; Wijedasa, L. S.; and Swinfield, T. 2016. The need for long-term remedies for Indonesia's forest fires. *Conservation Biology* 5–6.

Cortez, P.; and Morais, A. 2007. A Data Mining Approach to Predict Forest Fires using Meteorological Data. In *New Trends in Artificial Intelligence, 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence*, 512–523. APPIA.

Davies, D. K.; Ilavajhala, S.; Wong, M. M.; and Justice, C. O. 2009. Fire Information for Resource Management System: Archiving and Distributing MODIS Active Fire Data. *IEEE Transactions on Geoscience and Remote Sensing* 72–79.

Doerr, S. H.; and Santín, C. 2016. Global trends in wildfire and its impacts: perceptions versus realities in a changing world. *Philosophical Transactions of the Royal Society B: Biological Sciences* 20150345.

Doshi, J.; Garcia, D.; Massey, C.; Llueca, P.; Borensztein, N.; Baird, M.; Cook, M.; and Raj, D. 2019. FireNet: Real-time Segmentation of Fire Perimeter from Aerial Video. *arXiv preprint arXiv:1910.06407* .

Fogarty, L.; Pearce, H.; Catchpole, W.; and Alexander, M. 1998. Adoption vs. adaptation: lessons from applying the Canadian forest fire danger rating system in New Zealand. In *Proceedings, 3rd International Conference on Forest Fire Research and 14th Fire and Forest Meteorology Conference, Luso, Coimbra, Portugal*, 1011–1028.

Fukushima, K. 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics* 193–202.

Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; and Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 18–27.

Hochreiter, S.; and Schmidhuber, J. 1997. Long Short-Term Memory. *Neural Comput.* 1735–1780.

Huijnen, V.; Wooster, M.; Kaiser, J.; Gaveau, D.; Flemming, J.; Parrington, M.; Inness, A.; Murdiyarso, D.; Main, B.; and van Weele, M. 2016. Fire carbon emissions over maritime southeast Asia in 2015 largest since 1997. *Scientific Reports* 26886.

Jaafar, Z.; and Loh, T.-L. 2014. Linking land, air and sea: potential impacts of biomass burning and the resultant haze on marine ecosystems of Southeast Asia. *Global Change Biology* 2701–2707.

Karl Hille, N. 2015. Suomi NPP. URL https://www.nasa.gov/mission_pages/NPP/mission_overview/index.html. Date last accessed: 2021-03-16.

Kloprogge, P.; van der Sluijs, J. P.; and Petersen, A. C. 2011. A method for the analysis of assumptions in model-based environmental assessments. *Environmental Modelling & Software* 289–301. Thematic issue on the assessment and evaluation of environmental models and software.

Lohberger, S.; Stängel, M.; Atwood, E.; and Siegert, F. 2018. Spatial evaluation of Indonesia's 2015 fire affected area and estimated carbon emissions using Sentinel-1. *Global Change Biology* 644–654.

López, A. S.; San-Miguel-Ayanz, J.; and Burgan, R. E. 2002. Integration of satellite sensor data, fuel type maps and meteorological observations for evaluation of forest fire risk at the pan-European scale. *International Journal of Remote Sensing* 2713–2719.

MA, F. 2004. FARSITE: Fire Area Simulator–model development and evaluation. *USDA Forest Service, Rocky Mountain Research Station, Research Paper RMRS-RP-4 Revised.(Ogden, UT)* .

Maulana, S. I.; Syaufina, L.; Prasetyo, L. B.; and Aidi, M. N. 2019. Spatial Logistic Regression Models for Predicting Peatland Fire in Bengkalis Regency, Indonesia. *Journal of Sustainability Science and Management* 55–66.

National Aeronautics and Space Administration (NASA). 2021a. Aqua Earth-observing satellite mission — Aqua Project Science. URL https://aqua.nasa.gov. Date last accessed: 2021-03-16.

National Aeronautics and Space Administration (NASA). 2021b. Land, Atmosphere Near real-time Capability for EOS (LANCE). URL https://earthdata.nasa.gov/earth-observation-data/near-real-time. Date last accessed: 2021-03-16.

National Aeronautics and Space Administration (NASA). 2021c. Moderate Resolution Imaging Spectroradiometer (MODIS). URL https://modis.gsfc.nasa.gov. Date last accessed: 2021-03-16.

National Aeronautics and Space Administration (NASA). 2021d. Terra — The EOS Flagship. URL https://terra.nasa.gov. Date last accessed: 2021-03-16.

National Aeronautics and Space Administration (NASA). 2021e. Visible Infrared Imaging Radiometer Suite (VI-IRS). URL https://earthdata.nasa.gov/earth-observation-data/near-real-time/download-nrt-data/viirs-nrt. Date last accessed: 2021-03-16.

Page, S.; Hoscilo, A.; Langner, A.; Tansey, K.; Siegert, F.; Limin, S.; and Rieley, J. 2009. *Tropical peatland fires in Southeast Asia*, 263–287. Springer Berlin Heidelberg.

Page, S.; and Rieley, J. 2016. *Tropical Peat Swamp Forests of Southeast Asia*, 1–9. Springer Netherlands.

Page, S. E.; Rieley, J. O.; and Banks, C. J. 2011. Global and regional importance of the tropical peatland carbon pool. *Global Change Biology* 798–818.

Radke, D.; Hessler, A.; and Ellsworth, D. 2019. FireCast: Leveraging Deep Learning to Predict Wildfire Spread. In *Proc. of IJCAI*, 4575–4581. International Joint Conferences on Artificial Intelligence Organization.

San-Miguel-Ayanz, J.; et al. 2012. Comprehensive Monitoring of Wildfires in Europe: The European Forest Fire Information System (EFFIS). In *Approaches to Managing Disaster*, 87–108. IntechOpen.

Stocks, B. J.; Lynham, T. J.; Lawson, B. D.; Alexander, M. E.; Wagner, C. E. V.; McAlpine, R. S.; and Dubé, D. E. 1989. The Canadian Forest Fire Danger Rating System: An Overview. *The Forestry Chronicle* 450–457.

Stockwell, C. E.; Jayarathne, T.; Cochrane, M. A.; Ryan, K. C.; Putra, E. I.; Saharjo, B. H.; Nurhayati, A. D.; Albar, I.; Blake, D. R.; Simpson, I. J.; Stone, E. A.; and Yokelson, R. J. 2016. Field measurements of trace gases and aerosols emitted by peat fires in Central Kalimantan, Indonesia, during the 2015 El Niño. *Atmospheric Chemistry and Physics* 11711–11732.

U.S. Geological Survey. 2019. *Landsat 7 (L7) Data Users Handbook*. U.S. Geological Survey, 2nd edition.

Wagner, C. E. V. 1987. Development and Structure of the Canadian Forest Fire Weather Index System. *Can. For. Serv., Forestry Tech. Rep* .

Warren, M.; Hergoualc'h, K.; Kauffman, J.; Murdiyarso, D.; and Kolka, R. 2017. An appraisal of Indonesia's immense peat carbon stock using national peatland maps: Uncertainties and potential losses from conversion. *Carbon Balance and Management* 12.

Wijayanto, A. K.; Sani, O.; Kartika, N. D.; and Herdiyeni, Y. 2017. Classification Model for Forest Fire Hotspot Occurrences Prediction Using ANFIS Algorithm. *IOP Conference Series: Earth and Environmental Science* 012059.

World Bank. 2016. The cost of fire : an economic analysis of Indonesia's 2015 fire crisis (English). *Indonesia sustainable landscapes knowledge; note no. 1.* .