Predicting forest fires in Indonesia using remote sensing

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

[1]National University of Singapore

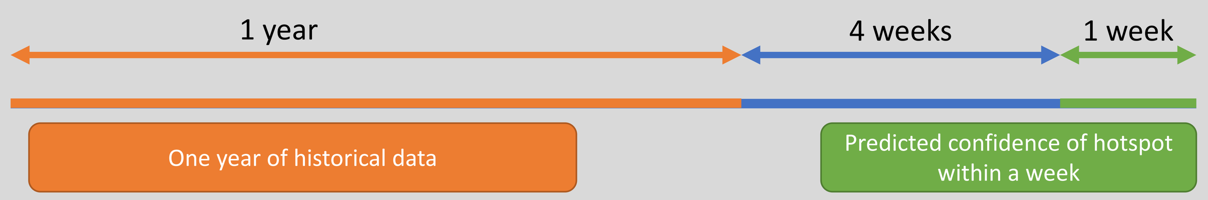
Introduction

In this work, we want to predict forest fires in Indonesia.

We use historical data to predict future forest fires. Specifically we formulate our approach as follows:

- Use 1 year of historical data
- Prediction for forest fire 4 weeks into future

The timeline is illustrated as follows:



We explore using a machine learning based approach to predict for forest fires

The approach is a supervised learning approach, we need both the data and the correct answers (labels) to train the machine learning model

We show that the approach is potentially a feasible approach to predicting forest fires

Motivation

- Recent advances in machine learning
- Wide range of possible applications for machine learning
- Remote sensing data are readily available
- Alternative to existing forest fire prediction systems

Current forest fire forecasting system:

- Uses handcrafted features
- Uses data from instruments on the ground location
- Requires a lot of domain expertise to build the system

Data and label

In the supervised learning formulation we require data and label:

Data – we use Landsat 7 satellite images, converted to histogram form

Label – we use FIRMS hotspot dataset as labels (desired prediction)

Model architecture

We used a convolutional LSTM neural network architecture. The convolution layer is used to capture relationship between the different bands of the satellite image. The long short term memory (LSTM) layer is used to capture temporal relationship between the historical data.

Experiment design

We evaluated the model to predict forest fires from June to September 2019. We used the area under the receiver operator characteristic(ROC) curve as a metric to evaluate our model.

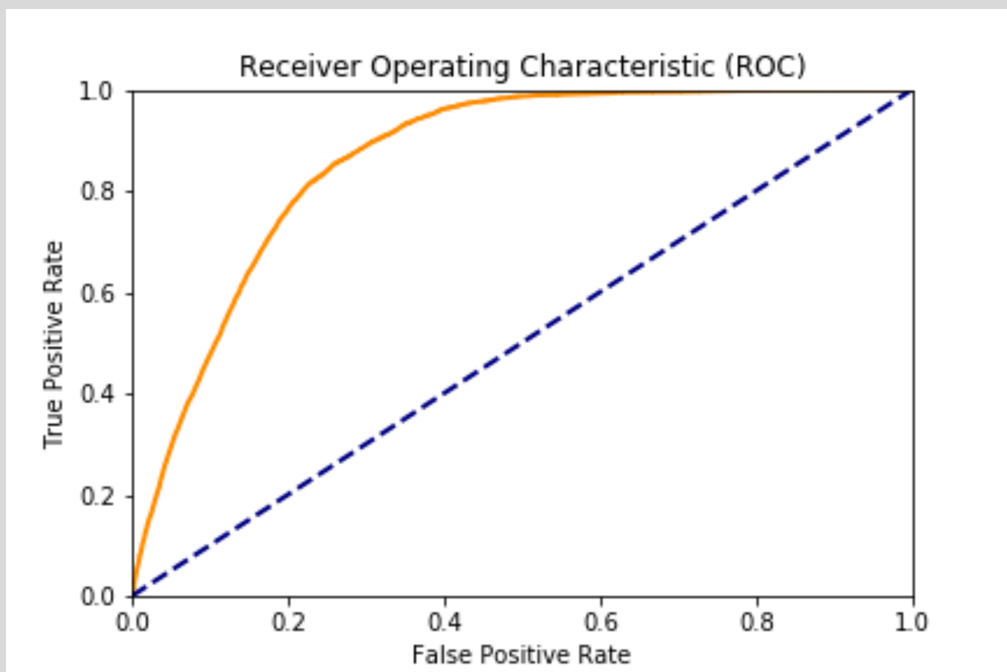
Area under receiver operating characteristic curve:

Y-axis : True positive rate (TPR)
X-axis : False positive rate (FPR)

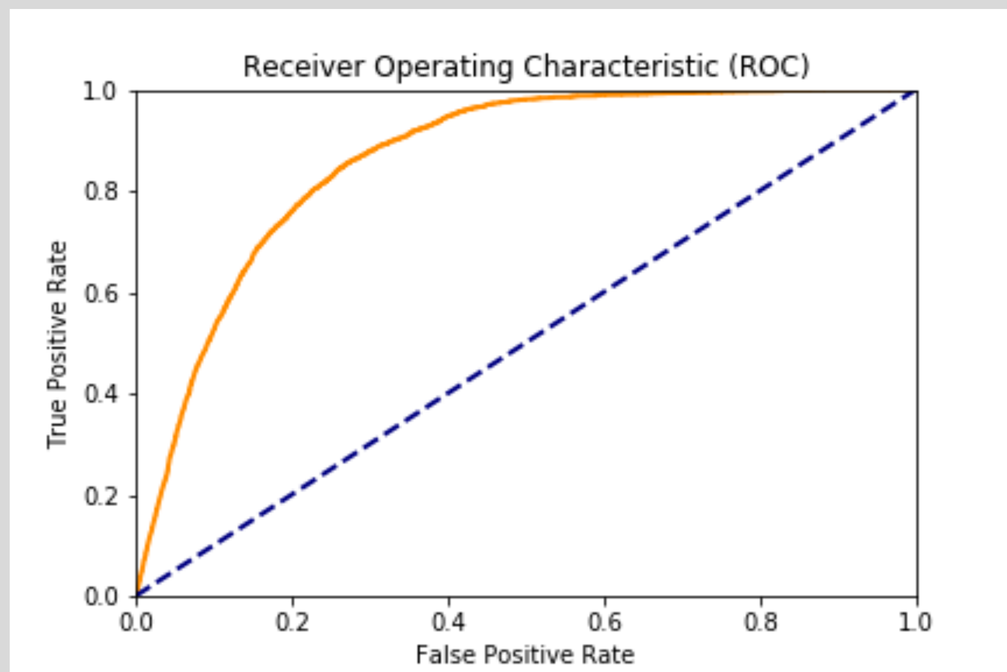
$$TPR = \frac{Number\ of\ true\ positive}{Number\ of\ true\ positive\ +\ Number\ of\ false\ negative}$$

$$FPR = \frac{Number\ of\ false\ positive}{Number\ of\ false\ positive\ +\ Number\ of\ true\ negative}$$

Experiment ROC curve plots



September 2019 hotspot
ROC curve, AUC = 0.8626



August 2019 hotspot
ROC curve, AUC = 0.8624

Experiment receiver operator characteristic

Hotspot Month	~ 3 Month	~ 6 Month	~ 9 Month	1 Year
June 2019	0.8438	0.8236	0.8568	0.8818
July 2019	0.8634	0.8494	0.8738	0.8867
August 2019	0.8275	0.8218	0.8437	0.8624
September 2019	0.8181	0.8234	0.8308	0.8626

Discussion

- Our initial exploration of this idea shows some promise
- This approach could be more cost-efficient → Using readily available remote sensing data VS maintaining instruments on the ground
- Further explorations could include:
  - Using multiple forms of remote sensing data
  - Analyzing underlying causes for prediction