

Time-Series Model for Climatological Forest Fire Prediction over Borneo

Arnida Lailatul Latifah^{ac2}, Furqon Hensan Muttaqien^{bc1}, Inna Syafarina^{c3}, Intan Nuni Wahyuni^{c4}

^aSchool of Computing, Faculty of Informatics, Telkom University
Jl. Telekomunikasi No. 1, Bandung, Indonesia

^bFaculty of Computer Science, Universitas Indonesia
Kampus UI, Depok, 16424, Indonesia
¹furqon.hensan@ui.ac.id (Corresponding author)

^cResearch Center for Computing, National Research and Innovation Agency
Jl. Raya Jakarta-Bogor KM. 47, Bogor, Indonesia
²arnida.l.latifah@brin.go.id
³inna002@brin.go.id
⁴inta008@brin.go.id

Abstract

Areas covered by tropical forests, such as Borneo, are vulnerable to fires. Previous studies have shown that climate data is one of the critical factors affecting forest fire. This study aims to predict the forest fire over Borneo by considering the temporal aspects of the climate data. A time series-based model, Long Short-Term Memory (LSTM), is used. Three LSTM models are applied: Basic LSTM, Bidirectional LSTM, and Stacked LSTM. Three different experiments from January 1998 to December 2015 are conducted by examining climate data, Oceanic Nino Index (ONI), and Indian Ocean Dipole (IOD) index. The proposed model is evaluated by Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and correlation number. As a result, all models can capture the spatial and temporal pattern of the forest fires for all three experiments, in which the best prediction occurs in September with a spatial correlation of more than 0.75. Based on the evaluation metrics, Stacked LSTM in Experiment 1 is slightly superior, with the highest annual pattern correlation (0.89) and lowest error (MAE= 0.71 and RMSE=1.32). This finding reveals that an additional ONI and IOD index as the prediction features would not improve the model performance generally, but it specifically improves the extreme event value.

Keywords: forest fire, climate, time-series data, Borneo, LSTM

1. Introduction

Forest and land fires are natural disasters that occur every year in Indonesia. Forest fires in Indonesia have happened since the 1970s, causing haze. The worst forest fire occurred in Sumatra and Borneo. According to the Indonesian Ministry of Environment and Forestry, forest fires have caused Borneo Island to lose more than 1 million ha of forest area in the last five years [1].

Forest fires can be caused by human activities, natural phenomena such as lightning, or a combination of both [2]. About 90% of forest fires in Indonesia are caused by human activities [3], intentional or unintentional. An example of intentional activity is clearing land for plantations using fire, while the unintended action is throwing cigarette butts carelessly. Both human activities and natural phenomena are hardly quantified. Besides, forest fires are strongly correlated to the dry season. In 2009, The Center for Research on the Epidemiology of Disaster (CRED) classified forest fires as a climatological disaster because their spread is closely associated with the dry season [2]. The most extensive fires occurred in 1997/1998, which coincided with the El

Niño Southern Oscillation (ENSO) drought phenomenon. The same incident happened again in 2014/2015 [2]. The study results in [4] also showed that the climate variable strongly correlates with fires in Borneo.

There are many studies about forest fires that implement machine learning, as described in [4], [5], [6]. The problem domains in those studies also vary, from fire detection to planning and policy. Some of the methods used for fire prediction are random forest [7], clustering [4], neural networks [6], and deep learning [8]. Both [6] and [8] use time-series data in their studies. They used long short-term memory (LSTM) as their initial model. LSTM is a neural network model developed to improve recurrent neural network (RNN) by implementing forget gate [9].

A recent study applied the Random Forest method to predict the burned area over Borneo based on temperature, relative humidity, precipitation, and wind speed [7]. Although their study gave promising results, it still inaccurately predicts extreme fires due to the ENSO phenomenon. In addition, their study also had not considered the predictor data as time-series data. Besides, [6] used three neural network models to solve the same problem in Alberta, Canada. The experimental results showed that LSTM performed better than RNN and back-propagation neural networks. Their study used maximum temperature, cooling degree day, heating degree days, rain total, snow total, snow on the ground, and the direction and speed of a maximum wind gust as the predictors' variables. Meanwhile, the target variables are the scale of forest wildfires determined by the fire's duration and the size of the burned area.

Unlike a study by [7] which predicted Borneo's forest fires by looking at the data without considering the time, this study aims to predict forest fires over Borneo based on climate variables time-series data. Implementation of a time-series model for predicting forest fire over Borneo is the novelty of this research. The result is expected to contribute to the decision-making of the stakeholders.

Numerous real-world applications have been studied using time-series data, such as speech recognition, object recognition in videos [10], handwriting [11], stock market prediction [12], or weather forecasting [13]. Some methods for time-series data analysis are tiled convolutional neural network, RNN, undecimated fully convolutional neural network (UFCNN), LSTM, and convolution and long short-term memory network (ConvLSTM) [9], [13], [14]. The newest model, called transformers, has proven to give better results for time-series data, as shown in [15] and [16].

Many studies show that the time-series model can give a promising result, so this study develops an LSTM model to predict forest fire. The choice of LSTM was motivated by [6] and [8], showing that LSTM is the most proven method used in forest fire prediction. The climate variables employed in the model are temperature, relative humidity, precipitation, and wind speed that affect the spread of forest fires as used in [7]. Furthermore, other variables that can describe the ENSO phenomenon are added, namely the Oceanic Niño Index (ONI) and the Indian Ocean Dipole (IOD) index.

2. Research Method

2.1. Data

The climate data are obtained from the ERA-Interim dataset [17]. The ERA-Interim dataset consists of 2-meter dew point temperature ($d2m$) in Kelvin, 2-meter temperature ($temp$) in Kelvin, and 10-meter wind speed (ws) in ms^{-1} . The dew point and 2-meter temperature variables are converted to Celcius and used to calculate the relative humidity ($rhum$) variable as implemented in [7]. Figure 1 describes the detailed conversion of temperature to relative humidity. The precipitation (pcp) variable uses the observational data from the Tropical Rainfall Measuring Mission (TRMM), which is available in the form of annual data over a period from 31 December 1997 to recent [18].

The ONI and IOD index data are derived from The National Oceanic and Atmospheric Administration (NOAA). The burned area dataset is collected from Global Fire Emission Database version 4 (GFED4) [19]. The GFED4 burned area is based on active fire detection from European Remote

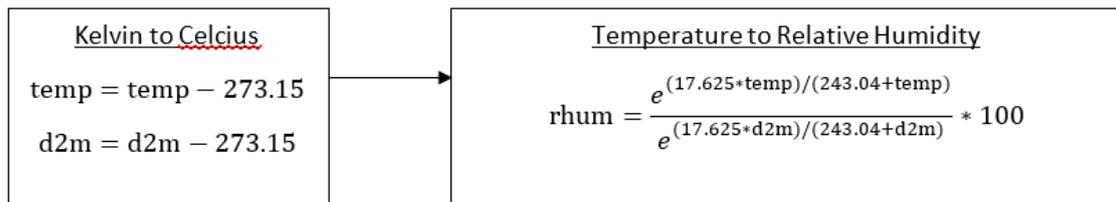


Figure 1. Conversion from temperature to relative humidity.

Sensing Satellite Along-Track Scanning Radiometer (ATSR) World Fire Atlas, Tropical Rainfall Measuring Mission Visible and Infrared Scanner (VIRS), and the Moderate Resolution Imaging Spectroradiometer (MODIS) burnt area product (MCD64A1) [19].

The data period used in this study is from January 1998 to December 2015 for all experiments. The domain area in this study is Borneo Island, which is covered by the latitude and longitude, 109.125°–118.875°E and 3.875°S–6.875°N. All climatology and burned area data are in 44 × 40 2-D array format. Each array element represents an area of 0.25° × 0.25°. While the ONI and the IOD index only have temporal information. Because the ONI and the IOD index do not have spatial information, both data are converted to a 2-D array format by setting the same value to all array elements.

In this study, each grid area represented as an array element is considered independent. Hence, there are 1760 sets of time-series data, as many as the number of array elements. Each set of time-series data consists of six predictors and one target in a daily format, which is joined into a single data frame and saved as a CSV file, as shown in Figure 2. So in total, there are 6573 × 1760 record data used for all experiments.

pcp	rhum	temp	ws	ONI	IOD	burned_area
14.053101539611800	88.92082313080640	26.521114238544400	1.8061947749798800	2.2	0.59	0.005249275825917720
13.297601699829100	89.15825472797600	26.64139086831730	1.8135989563113800	1.9	502.0	0.004013216122984890
11.526118278503400	89.4200274019031	26.741273243224700	1.5917123609254000	1.4	45.0	0.007319086231291290
12.017666816711400	89.4824900612019	26.85506329058750	1.4003978691017500	1.0	122.0	0.0067003704607486700
11.085970878601100	87.38661075627760	27.428960920765200	1.6900821986153600	0.5	219.0	0.0010938680497929500
7.637757301330570	85.62349123650390	27.010741268428900	1.6278461306491800	-0.1	261.0	0.000644950196146965
6.399472236633300	86.38263140009280	26.363402332320300	1.5271219680197200	-0.8	-288.0	0.0045324042439460800
16.052047729492200	87.41933362638620	25.960299715261600	1.6058041899111600	-1.1	-467.0	0.0017165378667414200
11.432831764221200	86.31028767040710	26.29512830390260	1.7524888698380200	-1.3	-0.39	0.00043294247006997500

Figure 2. Samples of time series data used in this study consist of precipitation (pcp), relative humidity (rhum), temperature (temp), wind speed (ws), ONI index, IOD index (IOD), and burned area.

2.2. Model and Experiment

Three types of LSTM models used in this study are Basic LSTM [9], Bidirectional LSTM (Bi-LSTM) [20], and Stacked LSTM [21]. Both bidirectional and stacked LSTM are improvements of basic LSTM and have been claimed to be more accurate for certain cases. All models are used to compare and determine the best model for the case in this study. The illustration of the architecture of Basic LSTM, Bidirectional LSTM, and Stacked LSTM can be seen in Figure 3.

In this study, three experiments based on the predictors' variables, as presented in Table 1, are conducted in which the burned area is the target variable. Each experiment's data is split into 16 years for training and two years for testing. The testing years are 2014 as a normal year and 2015

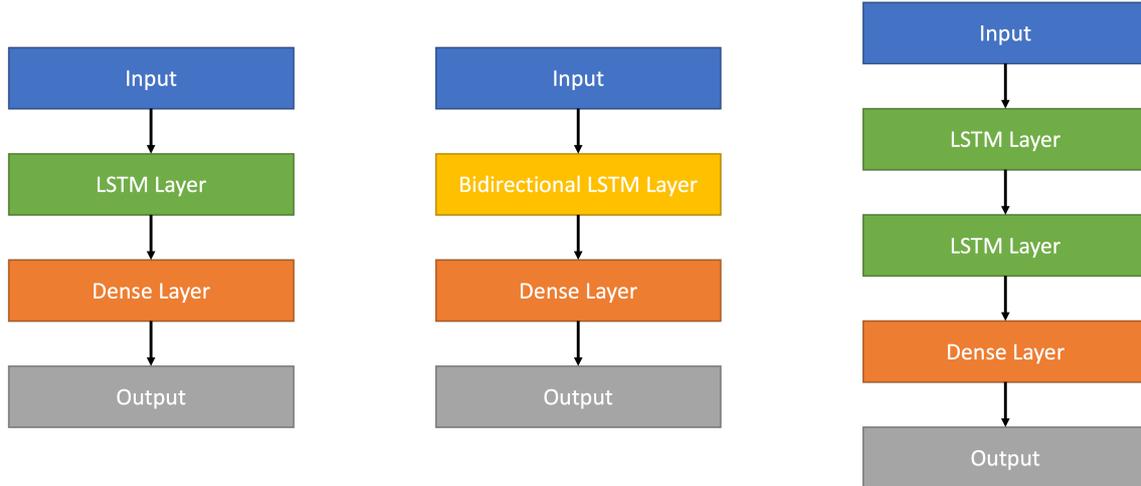


Figure 3. Basic LSTM (left), Bidirectional LSTM (middle), and Stacked LSTM architecture (right).

as an extreme year when El Nino occurred quite strongly. One hundred epochs are used in every experiment.

All three models are implemented in every experiment. Each model's hyperparameters are tuned in the number of hidden neurons and the activation functions of the dense layer until the best results are achieved. The activation functions used in the tuning process are Rectified Linear Unit (ReLU) and sigmoid. Sigmoid is the default activation function for time-series models such as LSTM [9], [6], while ReLU is more effective in deep networks [22]. These processes are performed on all 1760 grids individually so that each model can predict forest fires throughout Borneo.

Table 1. Experiments conducted in this study

Experiment	Predictor Variables
1	climate data (temp, rhum, pcp, ws)
2	climate data, ONI
3	climate data, ONI and IOD index

2.3. Evaluation Method

The mean absolute error (MAE) and the root mean square error (RMSE) are used to evaluate the model's accuracy. Both methods are chosen as they are the most common metric evaluation for a regression model. The equation for calculating MAE and RMSE can be seen in Equation 1 and Equation 2, respectively.

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - x_j| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - x_j)^2} \quad (2)$$

where y_j is the predicted value, x_j is the ground truth, and n is the number of data.

Additionally, Pearson correlation is used to measure the pattern similarity of the prediction and the ground truth in both spatial and annual patterns. The formula of Pearson correlation is presented in Equation 3 where \bar{y} and \bar{x} are the mean of the predicted value and the ground truth, respectively.

$$Corr = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

3. Result and Discussion

3.1. Spatial variability of burned area

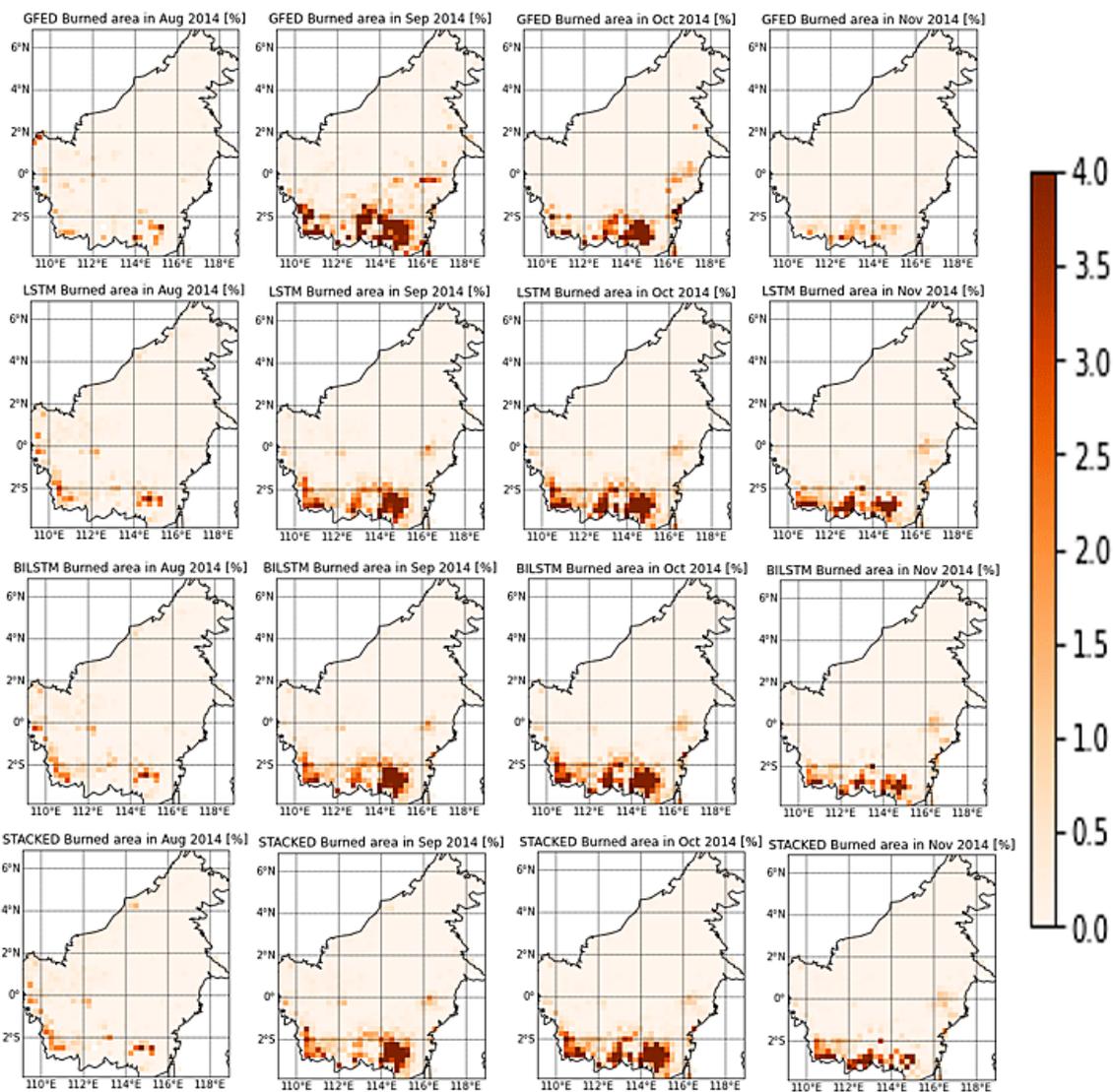


Figure 4. Spatial pattern of the burned area in August-November 2014. In row order: Ground truth, LSTM, Bi-LSTM, and stacked LSTM.

The annual forest fire over Borneo commonly occurs during the dry season. Then this section presents the model prediction result of the spatial variability of the burned area only from August until November for 2014 and 2015, see Figure 4 and 5. Both figures show the model prediction with Experiment 1 only as the results of the spatial pattern of Experiment 1-3 do not differ significantly. From the spatial pattern presented in Figure 4 and 5, the difference performance of all models are hardly observed. The model predictions agree with the ground truth (top panels Figure 4) in predicting the pattern of the forest fire in 2014, except in November when the mod-

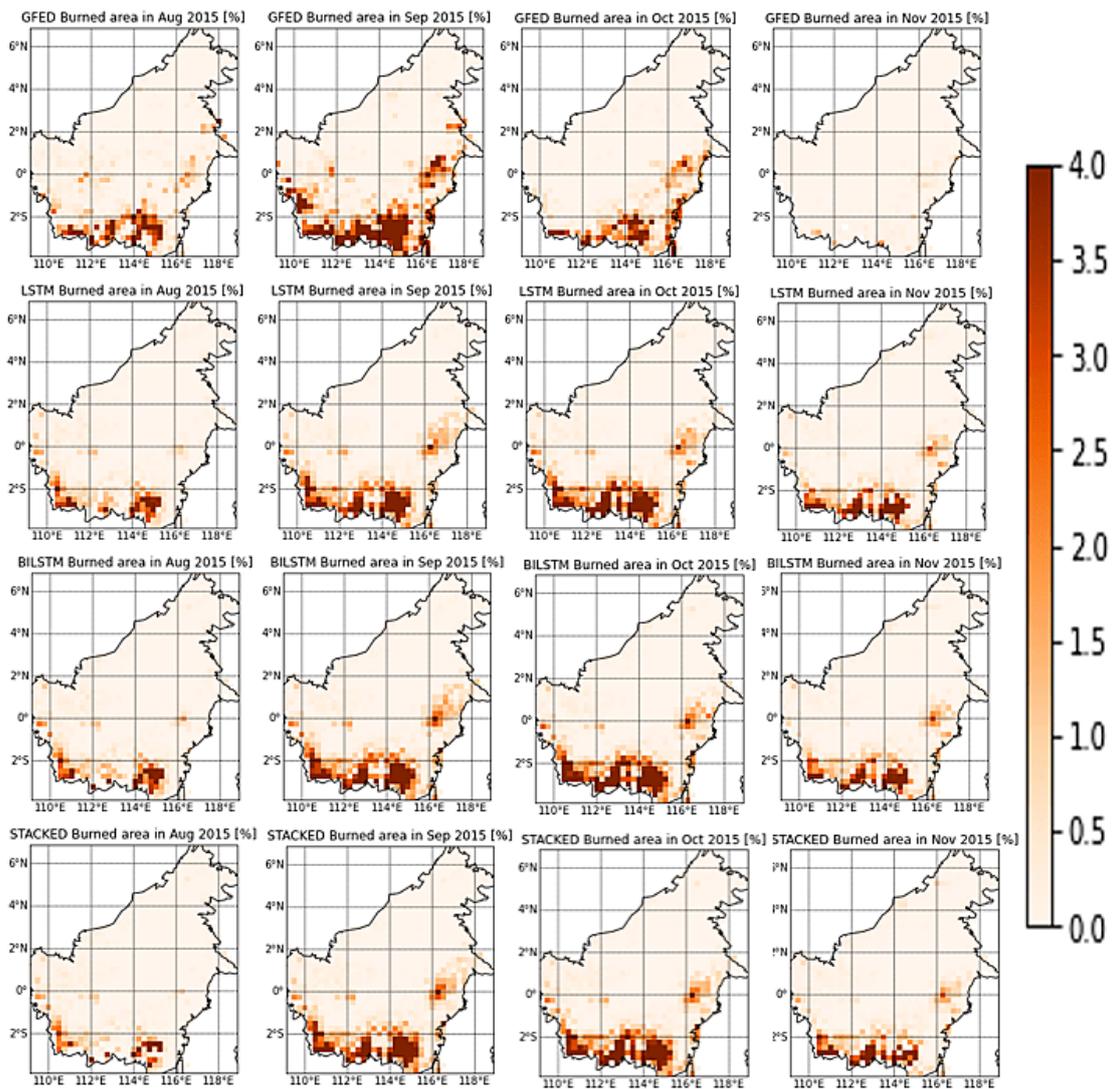


Figure 5. Spatial pattern of the burned area in August-November 2015. In row order: Ground truth, LSTM, Bi-LSTM, and stacked LSTM.

els predicted the fires still occurred largely in the southern part of Borneo while the ground truth shows only small fires.

Different from the 2014 event, the forest fire over Borneo in 2015 seemingly started earlier. The fire was already relatively high in August and getting more prominent in September and October. The models can also capture this early event, with a slight underestimation in August 2015. On the other hand, when the fire vanished in November 2015, the models overestimated the southern part of Borneo. This shows that all models overestimate the fire in November during the normal year (2014) or extreme year (2015).

Insignificant differences among the models are also observed from the evaluation metrics shown in Table 2. From August until November, the highest correlation is in September for all models and experiments, while the MAE and RMSE are also larger in September as the fire gets bigger. Comparing Experiments 1-3, the correlation is not significantly different, but the MAE and RMSE seem to increase by adding more predictor variables.

Table 2. Evaluation metrics of the spatial pattern of the burned area (Prediction vs Ground truth): RMSE(1×10^{-4}), MAE(1×10^{-4}), Correlation (%)

EXP Model	Aug			Sep			Oct			Nov		
	RMSE	MAE	Corr	RMSE	MAE	Corr	RMSE	MAE	Corr	RMSE	MAE	Corr
1 LSTM	56	13	56.36	149	35	79.57	89	23	70.67	80	17	50.39
	57	13	55.28	144	35	79.87	101	24	70.93	70	15	44.14
	57	13	54.03	143	34	80.33	95	23	71.57	76	16	54.81
2 LSTM	59	14	51.50	144	36	79.56	101	27	72.75	92	23	53.48
	58	14	53.58	141	36	79.12	117	30	70.92	87	22	54.97
	57	14	52.70	139	35	79.86	117	30	72.61	94	23	56.51
3 LSTM	59	14	53.55	137	35	79.59	111	29	73.76	98	24	55.98
	64	15	49.85	143	36	75.75	129	32	72.11	89	21	50.73
	62	15	49.85	142	36	75.87	131	33	70.85	97	24	50.05

3.2. Annual pattern of the total burned area

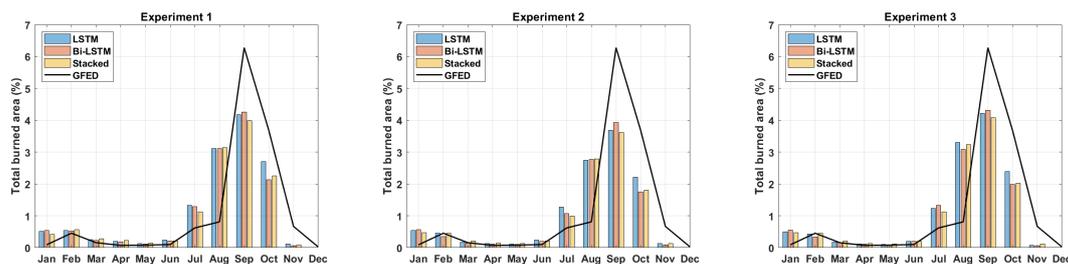


Figure 6. Annual pattern of the total burned area over Borneo (%) in 2014.

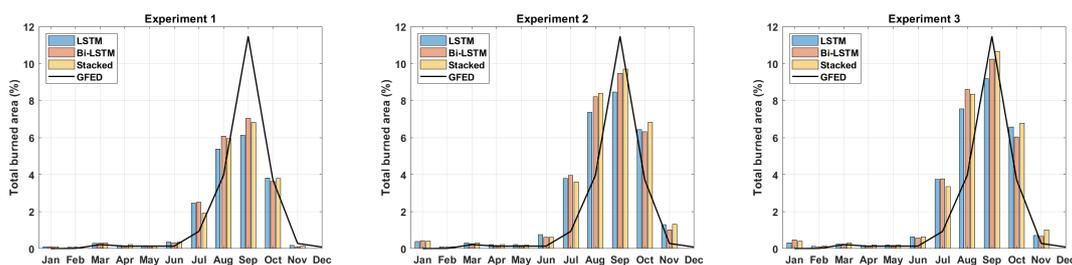


Figure 7. Annual pattern of the total burned area over Borneo (%) in 2015.

The annual pattern of the total burned area is shown in Figure 6 for 2014 and Figure 7 for 2015. To optimize the performance, the annual pattern of the model predictions is shifted one month earlier than the ground truth so that the forest fire’s peak matches the ground truth’s peak. The results show that the total burned area of all three LSTM models underestimates during the largest peak of forest fire in September - October 2014 for all three experiments. While, in the same period for 2015, only in September 2015, all three models show a slight underestimate for all three experiments and an overestimate for October - November 2015 for Experiments 2 and 3. It can

Table 3. Evaluation metrics of the annual pattern of the total burned area (Prediction vs Ground truth): **Correlation**, MAE, **RMSE**

Model	Experiment 1	Experiment 2	Experiment 3
LSTM	0.88	0.84	0.86
	0.72	0.97	0.90
	1.39	1.49	1.44
Bi-LSTM	0.88	0.85	0.87
	0.73	0.95	0.89
	1.31	1.52	1.48
Stacked	0.89	0.84	0.87
	0.71	0.97	0.89
	1.32	1.56	1.48

be seen that only Experiments 2 and 3 in 2015 can capture a similar result of the total burned area compared to observation data (GFED), even though there is an overestimate of Experiment 2 and 3 for June - August and October - November 2015. This underestimation is in line with the study by [7]. Nevertheless, their study cannot predict the most extreme fire in September 2015, while our study shows a good fit in Experiment 3.

Table 3 describes the error and correlation values for the two years' total burned area in Borneo. The evaluation metrics show that the best model performance in which the highest correlation (0.89), lowest MAE (0.71), and RMSE (1.32) is Stacked LSTM for Experiment 1. The models produced similar annual patterns with strong correlation values for all experiments. However, the error is high because the total burned area value cannot be achieved by the prediction results of all models, especially at the peak of the fires in September.

3.3. Time series prediction at local point

For the sake of brevity, only one local point from the third experiment was selected to show the performance of the LSTM model in predicting forest fire. Figure 8 shows the results.

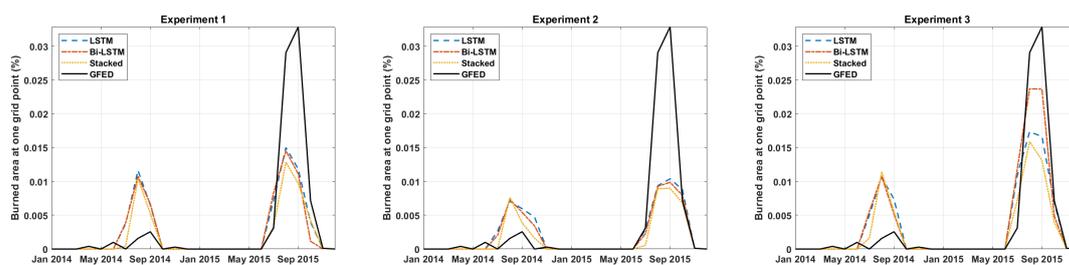


Figure 8. Time series of the burned area (%) in the testing period on a local point.

Figure 8 describes the time series of the burned area of model performances for the testing period at a certain point. All experiments can capture the peak of forest fire in September 2014 and 2015. In September 2014, all model and experiment results overestimated the observation data. Meanwhile, in September 2015, all model results underestimated the observation. There is a slight difference in the model result in Experiment 1 (only considerate climate variables) among three different LSTM models (top left Figure). For the second experiment, all three models also show a similar simulation result (middle plot) as in the first experiment. In Experiment 2 in September 2014, all three model results show closer to the observation (GFED). This indicates that adding ONI to the climate parameter for predictor could improve the model performance at the local point. By adding the IOD index, the model results of Experiment 3 in September 2015 show much better than the observation data compared to Experiment 1 and 2 as it gives a higher

prediction. Basic LSTM and Bi-LSTM show a similar result, while Stacked LSTM shows higher and close to the observation. Thus, Stacked LSTM could capture the time series of the burned observation area better than Basic LSTM and Bi-LSTM in the extreme year 2015.

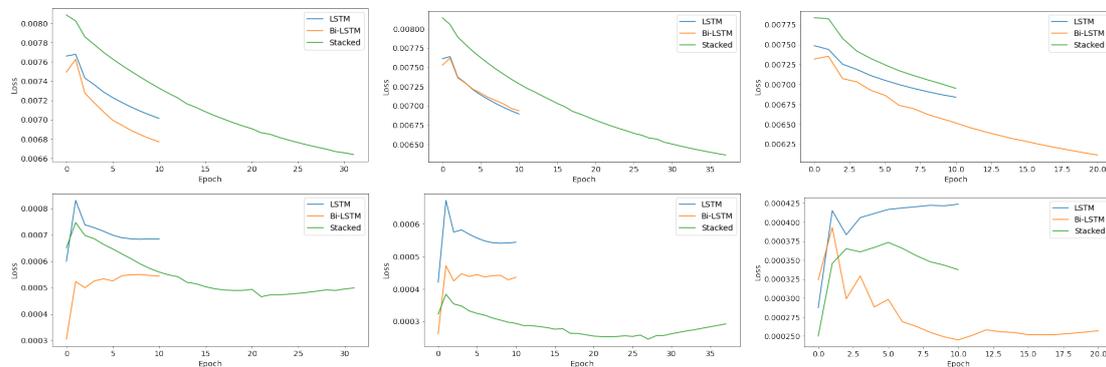


Figure 9. The loss function in the training (upper) and testing period (lower) related to Fig. 8. From left to right: Experiment 1, 2, 3.

The loss function for the training period (upper plots of Figure 9) shows a decrease. The first and second experiment shows a longer epoch for the Stacked LSTM model (> 30). Meanwhile, Bi-LSTM shows a longer epoch number for the third experiment, with about 20. This indicates that the calculation of Basic LSTM and Bi-LSTM in the first and second experiments reach an optimized result faster than Stacked LSTM for the training period. Meanwhile, the third experiment shows faster results in Basic LSTM and Stacked LSTM models. For the testing period (lower plots of Figure 9), the loss function shows an increase for the first epoch and then decreasing. The loss function for the testing period shows a similar trend to the training period.

In addition, to see the model's performance in the whole domain, the error distribution at all points is shown in Figure 10. The boxplots of the error emphasize our finding that all models perform similarly with approximately the same deviation. The difference is in the mean of the error distribution. Basic LSTM has a negative mean value, while other models have a primarily positive mean value.

4. Conclusion

Three different experiments have been conducted using three other LSTM models: Basic, Bidirectional, and Stacked. All models give similar results and predict the burned area over Borneo in 2014-2015 quite well. The high correlation between the spatial pattern predictions and the ground truth occurs in September and October, showing that the models give a good forecast for the burned area locations. However, the models show a significant overestimation in November. The annual pattern of all models' predictions strongly correlates with the ground truth. Nevertheless, Experiments 1-3 show some differences. There is a trend by adding a predictor variable; the peak of the total burned area seems to increase when ONI and IOD index is added. While the evaluation metrics show that Stacked LSTM in Experiment 1 performs best as it has the most considerable correlation and slightest error, the extreme fire occurring in September 2015 is only reached by Experiment 3. This describes that adding ONI and IOD index predict the burned area over Borneo higher, thus a better fit in September, but worse in other months (see the right panel in Figure 7). An improvement of the model prediction might be conducted by considering the spatial neighborhood of the burned area over Borneo. This study could be a good recommendation for policymakers to design an acceptable policy to prevent and control sites with more dangerous future fires.

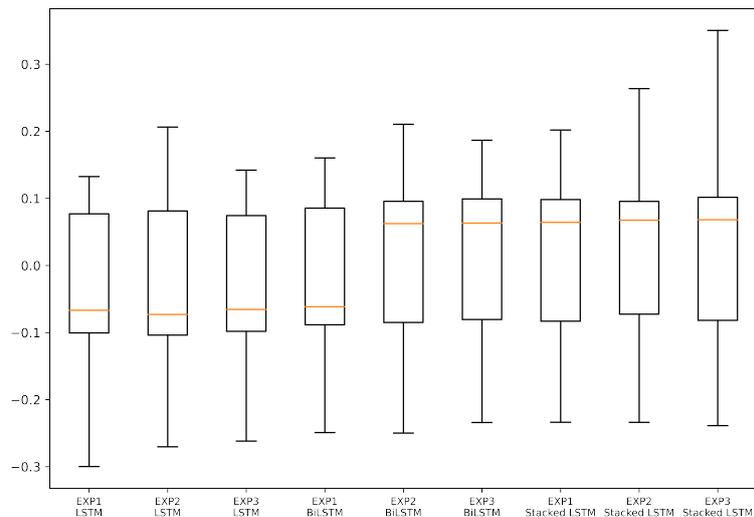


Figure 10. From left to right: Distribution of experiment bias using Basic LSTM, Bidirectional LSTM, and Stacked LSTM architecture, respectively

Acknowledgment

We want to thank Radityo Eko Prasajo, Ph.D. from Kata.ai & Universitas Indonesia, for his valuable and constructive suggestions during the planning and development of this research work. The computation in this work has been done using the facilities of MAHAMERU - BRIN HPC.

References

- [1] Sipongi. Luas kebakaran hutan dan lahan. [Online]. Available: <https://sipongi.menlhk.go.id/>
- [2] N. Yulianti, *Pengenalan Bencana Kebakaran dan Kabut Asap Lintas Batas*. Bogor: IPB Press, 2018.
- [3] E. Sumarga, "Spatial indicators for human activities may explain the 2015 fire hotspot distribution in central kalimantan indonesia," *Tropical Conservation Science*, vol. 10, p. 1940082917706168, 2017.
- [4] I. C. Hidayati, N. Nalaratih, A. Shabrina, I. N. Wahyuni, and A. L. Latifah, "Correlation of Climate Variability and Burned Area in Borneo using Clustering Methods," *Forest and Society*, vol. 4, no. 2, 7 2020.
- [5] P. Jain, S. C. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, and M. D. Flannigan, "A review of machine learning applications in wildfire science and management," pp. 478–505, 2020.
- [6] H. Liang, M. Zhang, and H. Wang, "A Neural Network Model for Wildfire Scale Prediction Using Meteorological Factors," *IEEE Access*, vol. 7, pp. 176 746–176 755, 2019.
- [7] A. L. Latifah, A. Shabrina, I. N. Wahyuni, and R. Sadikin, "Evaluation of Random Forest model for forest fire prediction based on climatology over Borneo," in *2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*. IEEE, 10 2019, pp. 4–8. [Online]. Available: <https://ieeexplore.ieee.org/document/8949588/>

- [8] Z. Li, Y. Huang, X. Li, and L. Xu, "Wildland Fire Burned Areas Prediction Using Long Short-Term Memory Neural Network with Attention Mechanism," *Fire Technology*, 2020.
- [9] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 11 1997.
- [10] C. Gonzalez Viejo, S. Fuentes, D. D. Torrico, and F. R. Dunshea, "Non-contact heart rate and blood pressure estimations from video analysis and machine learning modelling applied to food sensory responses: A case study for chocolate," *Sensors*, vol. 18, no. 6, p. 1802, 2018.
- [11] C. Taleb, M. Khachab, C. Mokbel, and L. Likforman-Sulem, "Visual representation of online handwriting time series for deep learning parkinson's disease detection," in *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, vol. 6. IEEE, 2019, pp. 25–30.
- [12] M. Wen, P. Li, L. Zhang, and Y. Chen, "Stock market trend prediction using high-order information of time series," *Ieee Access*, vol. 7, pp. 28 299–28 308, 2019.
- [13] J. C. B. Gamboa, "Deep learning for time-series analysis," *CoRR*, vol. abs/1701.01887, 2017. [Online]. Available: <http://arxiv.org/abs/1701.01887>
- [14] H. Lin, Y. Hua, L. Ma, and L. Chen, "Application of ConvLSTM network in numerical temperature prediction interpretation," in *ACM International Conference Proceeding Series*, vol. Part F1481, 2019, pp. 109–113.
- [15] N. Wu, B. Green, X. Ben, and S. O'Banion, "Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case," 1 2020. [Online]. Available: <http://arxiv.org/abs/2001.08317>
- [16] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang, and X. Yan, "Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting," 6 2019. [Online]. Available: <http://arxiv.org/abs/1907.00235>
- [17] European Centre for Medium-range Weather Forecast (ECMWF). (2011) The era-interim reanalysis dataset, copernicus climate change service (c3s). [Online]. Available: <https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/reanalysis-datasets/era-interim>
- [18] Tropical Rainfall Measuring Mission (TRMM). (2011) Rmm (tmpa) rainfall estimate l3 3 hour 0.25 degree x 0.25 degree v7. [Online]. Available: <http://dx.doi.org/10.5067/TRMM/TMPA/3H/7>
- [19] L. Giglio, J. T. Randerson, and G. R. van der Werf, "Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4)," *Journal of Geophysical Research: Biogeosciences*, vol. 118, no. 1, 3 2013.
- [20] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm networks," in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, vol. 4, 2005, pp. 2047–2052 vol. 4.
- [21] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," 2013. [Online]. Available: <https://arxiv.org/abs/1312.6026>
- [22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.