Journal of Natural Resources and Environmental Management
https://doi.org/10.29244/jpsl.15.2.255

RESEARCH ARTICLE

Application of Random Forest Algorithm to Analyze the Confidence Level of Forest Fire Hotspots in Riau Peatland

Mitra Unik[a,b], Imas Sukaesih Sitanggang[a], Lailan Syaufina[c], I Nengah Surati Jaya[d]

[a] Department of Computer Science, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, 16680, Indonesia
[b] Department of Informatics Engineering, Faculty of Computer Science, University of Muhammadiyah Riau, Pekanbaru, 28156, Indonesia
[c] Department of Silviculture, Faculty of Forestry and Environment, IPB University, Bogor, 16680, Indonesia
[d] Department of Forest Management, Faculty of Forestry and Environment, IPB University, Bogor, 16680, Indonesia

Article History
Received 2 November 2023
Revised 20 August 2024
Accepted 20 August 2024

Keywords
confidence analysis, forest fire, machine learning, peatland, Random Forest algorithm

ABSTRACT

Forest fires pose a significant challenge in Riau Province, Indonesia, especially in peatland areas. This study employs the Random Forest (RF) algorithm to analyze the confidence levels of hotspots, aiming to predict potential fire occurrences and improve fire management strategies. The research focuses on peatlands spanning 3.86 million ha, using key variables such as NDVI, surface temperature, and peat thickness derived from satellite data. The model achieved an average AUC of 0.732 and a classification accuracy of 70.3%, with medium-confidence hotspots demonstrating the best predictive performance (AUC: 0.707, F1-score: 0.804). However, the model struggled with low-confidence hotspots, reflecting challenges in distinguishing less prominent patterns in the data. Compared to other methods, RF demonstrates strong potential in handling complex environmental datasets, making it a valuable tool for hotspot prediction. This study contributes to understanding forest fire risks in peatlands and provides actionable insights for improving preparedness and mitigation efforts.

Introduction

Hotspot confidence is an important factor in forest fire preparedness, including resource procurement, resource relocation and preparation for fire suppression activities. This confidence value indicates the likelihood of a hotspot developing into a fire. It is calculated based on the geometric mean of five factors, such as temperature, cloud cover, smoke cover, vegetation and water conditions. These factors help detect hotspots using satellite data such as MODIS (Moderate Resolution Imaging Spectroradiometer) to map the relationship between hotspots and fire occurrence [1]. Hotspots usually indicate areas with higher temperatures than their surroundings, but not all hotspots can be confirmed as a sign of fire. Experts state that a hotspot can only be considered a fire indicator if it is detected repeatedly for two to five consecutive days. Indonesia has the second largest peatland in the world after Brazil, with 22.5 million ha [2] spread across Sumatra, Kalimantan, and Papua [3]. Peatland fires are a recurring problem, causing environmental and health damage [4,5]. In 2019, 1.6 million ha of land burned with 22,389 hotspots, mostly on peatlands [6,7]. These fires exacerbate global climate change [8] destroying biodiversity [9,10] and produced a haze that spread across the country [11].

The National Institute of Aeronautics and Space (LAPAN) of Indonesia has established a system to classify hotspots based on confidence intervals, which determine the necessary actions to address potential fire risks. Hotspots with a confidence level below 30% are categorized as low confidence, requiring monitoring but no immediate action. Confidence levels ranging between 30% and 80% fall into the medium confidence category, which warrants an alert and preparedness. Lastly, hotspots with confidence levels of 80% or higher are categorized as high confidence, demanding immediate countermeasures [12]. In 2019, Indonesia

Corresponding Author: Mitra Unik mitraunik@apps.ipb.ac.id Department of Computer Science, IPB University, Bogor, Indonesia.

recorded a total of 22,389 hotspots, marking an increase compared to 2018. However, this figure was significantly lower–about 45% less–than the 40,727 hotspots detected in 2015. This trend highlights the fluctuating nature of hotspot occurrences and underscores the need for effective management strategies to mitigate fire risks [13].

Interventions to reduce risks and improve forest and land fire preparedness can be achieved by predicting the occurrence of such fires [14]. Several studies have applied machine-learning approaches in this domain. For example, Murphy [15] introduced a conceptual framework modified to include examples relevant to fire and forest management. Stojanova et al. [16] evaluated various machine learning methods for fire prediction in Slovenia using geographic, remote sensing, and meteorological data. The methods evaluated included single classifiers, such as K-Nearest Neighbor (KNN), naïve Bayes, Decision Trees (DT) using J48 and jRIP algorithms, Logistic Regression (LR), Support Vector Machines (SVM), and Bayesian Networks (BN). Additionally, ensemble methods, such as AdaBoost, DT with bagging, and Random Forest (RF), were examined. The results indicated that ensemble methods, particularly DT with bagging and RF, provided the best performance. The bagging method had higher precision, while RF demonstrated better recall.

In recent studies, Yu et al. [17] applied the RF method to predict fire risks in Cambodia using publicly available remote sensing data. Additionally, the maximum entropy (MaxEnt) method was utilized to forecast fire occurrences. Another study by de Angelis et al. [18] utilized the maximum entropy (MaxEnt) method to predict daily fire risk in the mountainous region of Canton Ticino, Switzerland, incorporating various meteorological variables and fire indices. Dutta et al. [19] employed a two-stage approach combining an ensemble of unsupervised deep belief neural networks (DBNet) and conventional ensemble machine learning to predict weekly forest fire hotspot occurrences. The authors found that the bagging classifier and the conventional KNN classifier were the two most accurate models, achieving 94.5% and 91.8% accuracy.

In Indonesia, Shofiana and Sitanggang [20] analyzed hotspot data from NASA FIRMS for 2014–2015 in Sumatra and Kalimantan. Their study identified 484 hotspots using the SPADE algorithm, 58 of which had low confidence levels. However, only 21 hotspots were eligible for verification using Landsat-8 satellite images. The verification revealed that 85.71% of the hotspots experienced decreased confidence due to haze interference. Similarly, Nurpratami and Sitanggang [21] employed a spatial entropy-based DT algorithm to classify forest fire hotspots in Bengkalis Regency, Riau Province, using NASA FIRMS data from 2008. The DT achieved an average accuracy of 89.04% for the training set and 52.05% for the test set. This study generated 255 classification rules based on proximity to city centers, rivers, roads, income sources, land cover, population, rainfall, schools, temperature, and wind speed.

Various studies have been conducted on forest and land fire prediction using machine learning approaches, particularly RF algorithms [22]. However, no study has specifically addressed the analysis of the confidence level of the occurrence of forest and land fire hotspots in peatlands using the RF algorithm. The purpose of this research was to overcome the obstacles in analyzing the level of confidence in the occurrence of hotspots that have great potential to become actual fire points, as well as anticipatory steps in dealing with recurrent forest and land fires. This research is expected to help understand the role of confidence levels in identifying the likelihood of fire occurrence. The RF algorithm was used as the confidence analysis approach in this study, and three different dataset groupings were used as input variables: Area of Interest (AOI), hotspot distribution, and interpretation of land conditions in the Riau Province, Indonesia.

Materials and Methods

Location and Time of Research

The study area is in Riau Province, which has an area of 87,023.66 km^2 or 9,068,997 ha with a population of 6,493,603 people [23]. This study focused on a peat area of 3,863,759.76 ha, which is astronomically located at latitudes between 0.86° and 2.49° N and longitudes between 100.30° and 103.76° E (Figure 1). The research period was from January 1st to July 31th, 2019.

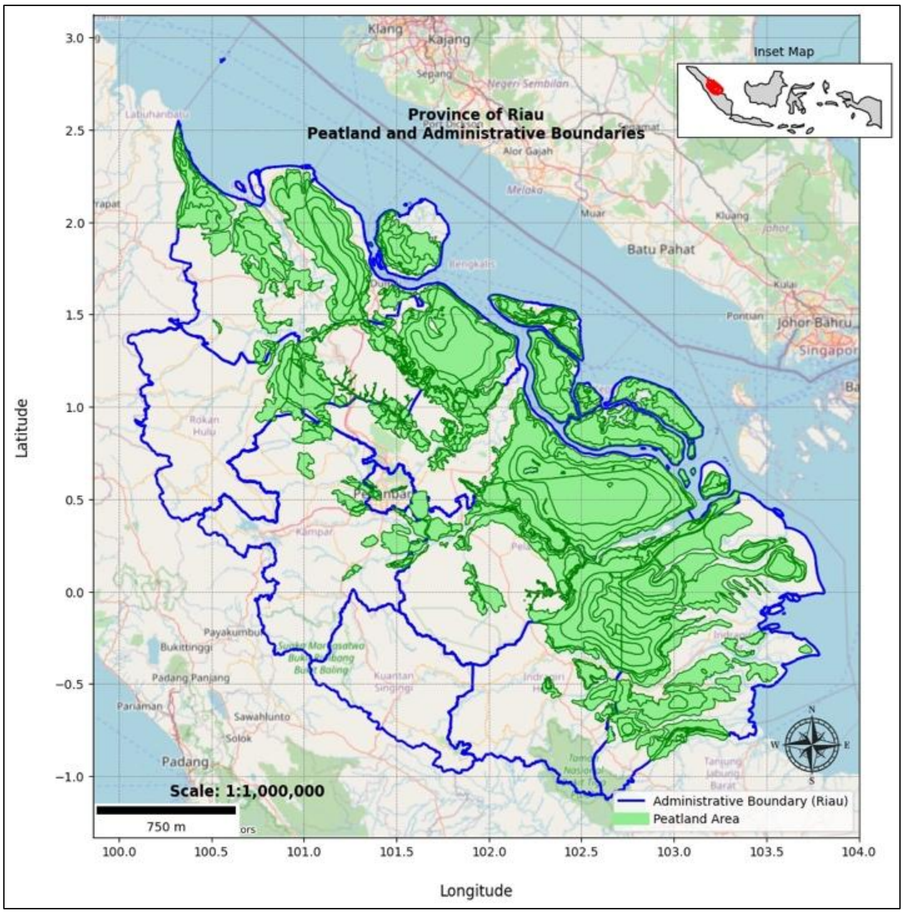


Figure 1. Study area in peatland of Riau Province, Indonesia.

Data Collection and Analysis Methods

This study used three different datasets as input variables: Area of Interest (AOI), hotspot distribution, and interpretation of land conditions. For each dataset, attributes related to peatland fires were selected. After obtaining the relevant attributes, all attributes were combined into one new dataset. The data used in this study were obtained from various sources, as described in detail in Table 1.

Table 1. Summary of research data used.

Variable	Unit	Temporal resolution	Spatial resolution	Resource
Area of Interest (AOI)				
Image map of Riau Province	km^2	8 days	1 km (0.928 km)	C6 MODIS LST MOD11A2 V6
Riau Province boundary	km^2	Unitless	Unitless	Geospatial Information Agency Indonesia (BIG)
Riau peat area boundary	km^2	Unitless	Unitless	BIG
Hotspots distribution				
Hotspots – NASA		Day		FIRMS NASA 2019
Hotspots – BRIN		Day		National Research and Innovation (BRIN) - Fire Hotspot 2019
Interpretation of land conditions				
NDVI	m^2	Month	30	Landsat 8 OLI – C2L2
Elevation	m	Month	0.27-arcsecond	National Digital Elevation Model (DEMNAS)
Peat thickness	m	Month	AOI	Centre for Research and Development of Agricultural Land Resources (BBSLDP)
Surface temperature (Temp)	^0C	Day	AOI	C6 MODIS LST MOD11A2 V6

Identify Hotspot Distribution and Area of Interest

Information on the distribution of hotspots within the AOI was represented in maps from satellite imagery data (Terra and Aqua C6 MODIS LST MOD11A2 V6) and by considering the AOI boundaries set by the BIG. A visual map of the AOI was created using geographic information system software, and the results are shown in Figure 2. Results of overlaying the distribution of selected hotspots with MODIS imagery in the AOI. The hotspot distribution is identified after data pre-processing, which involves several important steps, such as attribute selection, missing value filling, data merging, and distribution balancing for each class. All hotspot confidence attribute values from 0 to 100 were used without decimal rounding for the hotspot location coordinates. The hotspot dataset was created by selecting hotspot data from the peatlands and obtaining peat thickness information. This study used 1,419 datasets selected from 41,059 rows of data sets in 2019. Table 2 lists the hotspot confidence categories and frequency intervals before and after the selection.

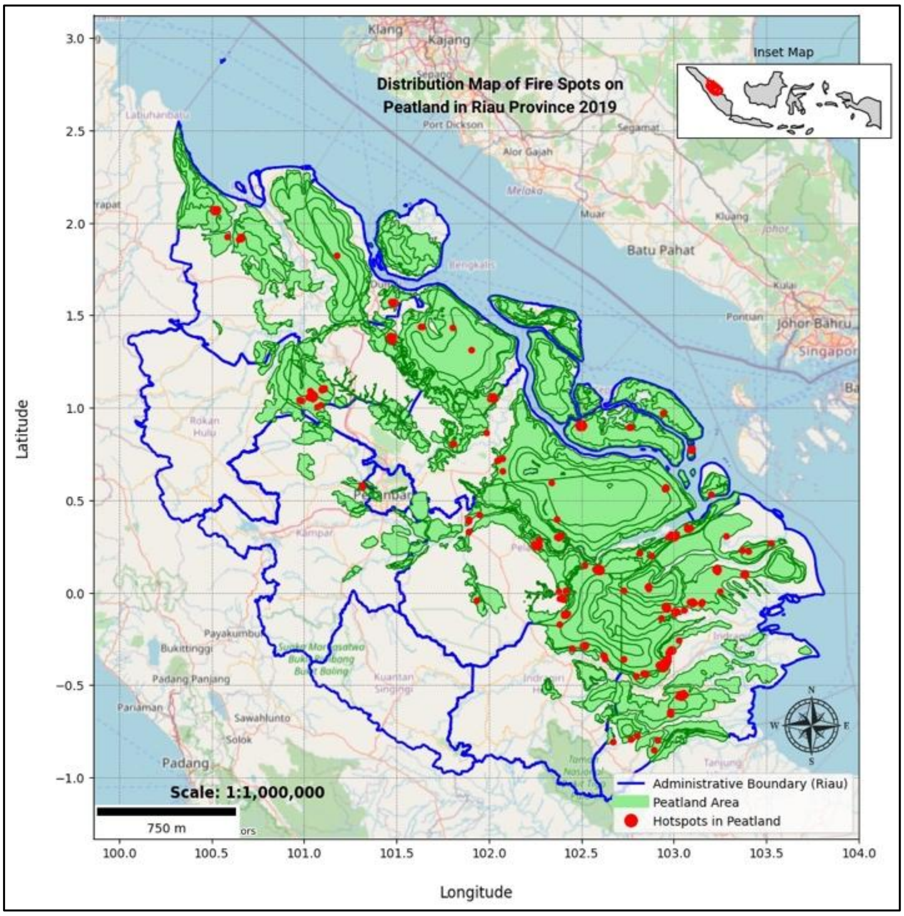


Figure 2. Results of overlaying the distribution of selected hotspots with MODIS imagery in the AOI area.

Table 2. Number of hotspots (before and after selection).

Hotspot resources	Number of hotspots		Confidence frequency		
	Before selection (Indonesia)	After selection (AOI)	Low	Medium	High
FIRM NASA	19,204	206			
BRIN	21,855	1,213			
Research hotspot datasets	41,059	1,419	121 (8.53%)	946 (66.67%)	352 (24.81%)

This research focuses on eight key attributes of hotspots. These include coordinate points (latitude and longitude), acquisition date (acq_date), and acquisition time (acq_time). Other attributes are soil type, peat thickness class, temperature class (Temp), Normalized Difference Vegetation Index (NDVI) class, and hotspot confidence level. These attributes are detailed in Table 3.

Table 3. Dataset attribute selection.

Attribute	Description
Latitude	Latitude coordinates of hotspot (°)
Longitude	Longitude coordinates of hotspot (°)
Brightness temperature	Temperature of hotspot in channel 21 and 22 (K)
Acq_Date	Date of fire incident
Acq_Time	Time of fire occurrence
Satellite	Satellite
Confidence	Hotspot quality (0–100%)
NDVI	Landsat 8 OLI-C2L2
Surface temperature	Temperature of hotspot in channel 31 (K)
Soil type	Soil characteristics
Peat thickness	The size of the peat layer in the soil

Interpretation of Land Conditions

The NDVI data, derived from Landsat 8 OLI-C2L2 satellite images, was used to represent vegetation land cover. Additional variables included Elevation from DEMNAS, peat thickness from BBSLDP, and surface temperature from C6 MODIS LST MOD11A2 satellite images. These variables were crucial in determining the characteristics of hotspot confidence levels. The classifications of NDVI, peat thickness, surface temperature, and soil type were further divided into specific classes with frequency counts, as detailed in Table 4.

Table 4. Interpretation of classification distribution of land condition dataset.

NDVI		Peat thickness		Surface temperature		Soil type	
Classify	Freq	Classify	Freq	Classify	Freq	Classify	Freq
Non-vegetation	2	Shallow	40	≤20	2	Fibric organosols	65
Sparse vegetation	27	Medium	291	>20–22	2	Hermic organosols	991
Medium vegetation	250	Deep	322	>22–24	4	Sapric organosols	363
Dense vegetation	561	Very deep	766	>24–26	5		
Very dense vegetation	579			>26–28	23		
				>28–30	68		
				>30–32	272		
				>32–34	752		
				>34–36	131		
				>36–38	87		
				>38–40	58		
				>40	15		

The land cover dataset was established using spatial operations from Landsat 8 OLI-C2L2 satellite data. Land cover value (NDVI) is a spectrum-based greenness index for measuring and monitoring plant growth, vegetation cover, and biomass production from multispectral satellite data [24]. NDVI is a combination of the red and NIR bands. The NDVI calculation is given by the Equation 1 [25].

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \tag{1}$$

where NIR is the near-infrared band (0.841–0.876 mm) and red is the red band (0.620–0.670 mm).

Pre-Processing Data

After completing the pre-processing steps, the dataset was finalized with 1,419 rows and nine columns. It includes a total of eight features, evenly divided into categorical and numerical types. Specifically, there are four categorical features and four numerical features. These features are detailed in Table 5. The Pre-processing was carried out in five stages, as shown in Figure 3. First, it was imported using the pandas library. The data used consist of the columns "*HSConf_Class,"* Soil Type, "Peat thickness," "NDVI," and "Temp.". Second, initialize *LabelEncoder* from the Scikit-Learn library and proceed to the stage, Encoding Categorical Variables, to convert each categorical variable into an integer. This stage is performed such that categorical variables can be used in classification models that require numerical data. Fourth, normalization or standardization was performed using *StandardScaler* from Scikit-Learn. This normalization ensured that the data had a mean of 0 and a standard deviation of 1. Finally, Train-Test Split. The data were split into training (train) and testing (test) data at a ratio of 80:20. Relationship visualization uses Seaborn to create pair plots that visualize the relationships between multiple numerical variables and differentiate them by class *i* as the target class, as shown in Figure 4.

Table 5. Example of a dataset table.

#	HSConf_ category	Soil type	Peat thickness	NDVI_Class	Surface_ Temperature	Acq_Date	Acq_TIME (WIB)
1	Low	Hemic organosol	Very deep	Dense vegetation	>30–32	2019-01-09	10:45:00
2	Low	Hemic organosol	Deep	Dense vegetation	>34–36	2019-02-06	11:07:00
…							
1419	High	Sapric organosol	Medium	Very dense vegetation	>32–34	2019-07-31	18:43

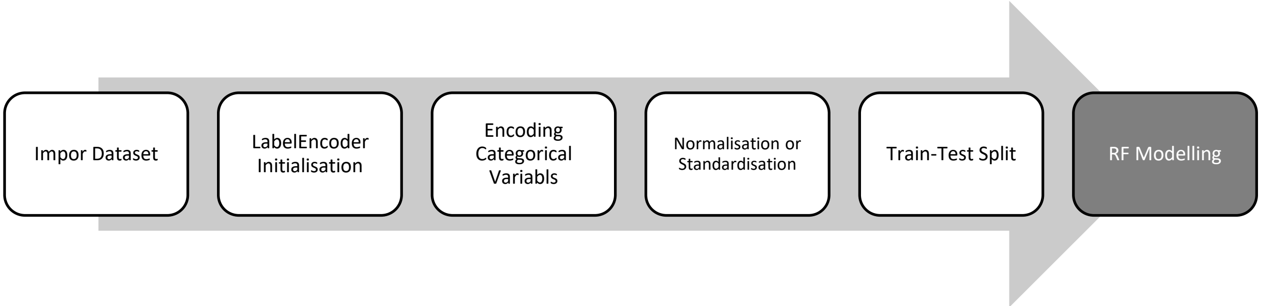


Figure 3. Pre-processing stage.

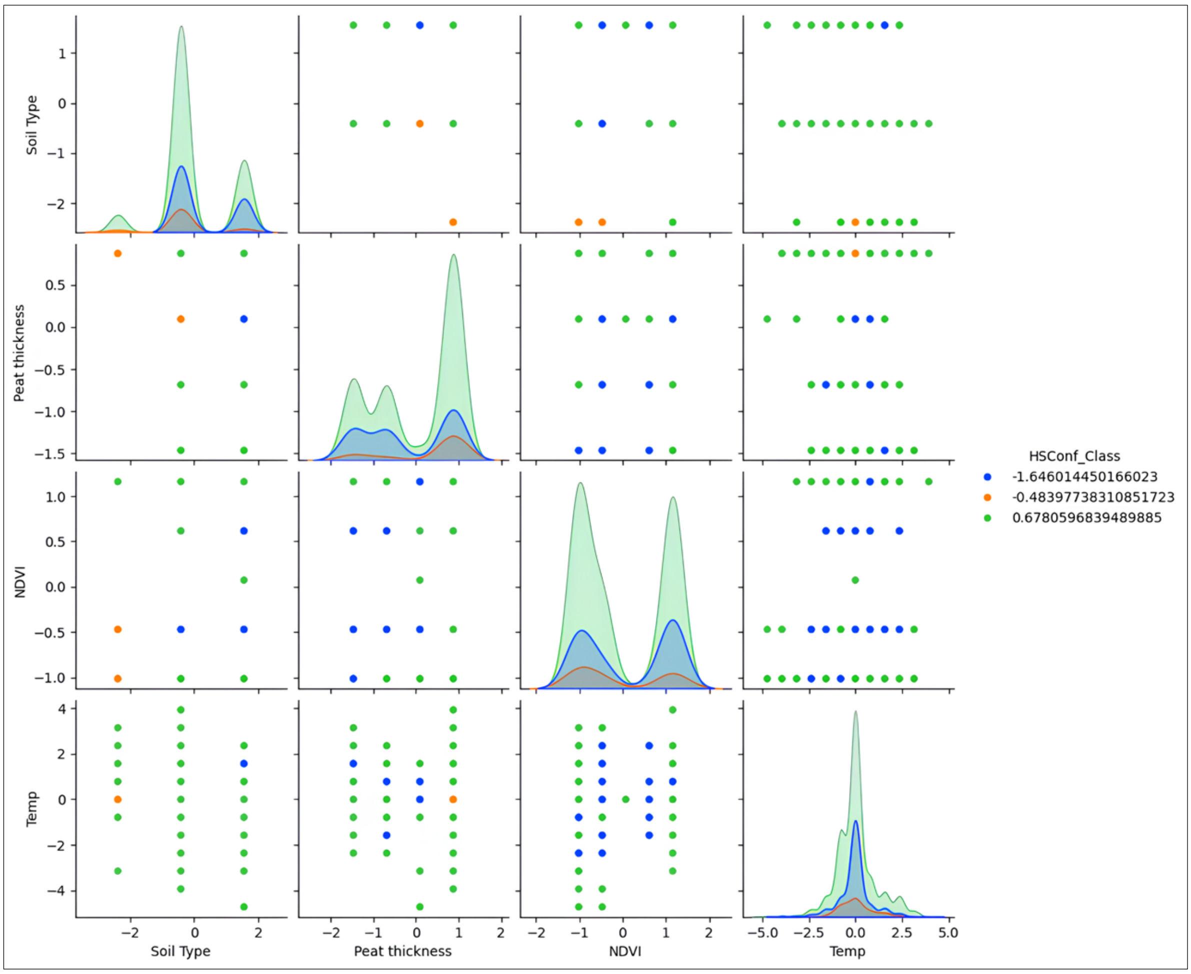


Figure 4. Seaborn's pair plots visualized the relationship between the numerical variables and target classes.

Seaborn's pair plot shows the relationship between the four variables of soil type, peat thickness, NDVI, and temperature, and the distribution of each variable. The colors in the scatter plot indicate three different classes of the HSConf_Class variable: blue (–1.646041450166023), orange (–0.48397738310851723), and green (0.6780596839489885). The soil type distribution showed a bimodal pattern, especially for the green class, indicating two main data groups for this variable. Similarly, the peat thickness distribution showed two major peaks, indicating a potential division of the data into two major groups. The NDVI distribution showed several peaks, indicating significant variation in the data, especially for the green and blue classes. The temperature distribution showed one main peak with a wide distribution, where the green class appeared to be dominant at various temperature values. Analysis of the relationships between the variables showed no strong correlation between the pairs of variables tested. For example, the relationship between soil type and peat thickness, NDVI and temperature did not show a clear correlation. Similarly, there was no significant relationship between peat thickness and NDVI and temperature, and between NDVI and temperature.

Results

Classification Modelling using Random Forest Algorithm

The classification model was constructed using the Random Forest algorithm with 500 DT in the ensemble. A 10-fold cross-validation method was implemented, with 80% of the dataset utilised for training and 20% for testing. This approach ensured a rigorous evaluation of the model's predictive performance and minimised the risk of overfitting. An illustration of the Pythagorean Forest is shown in Figure 5.

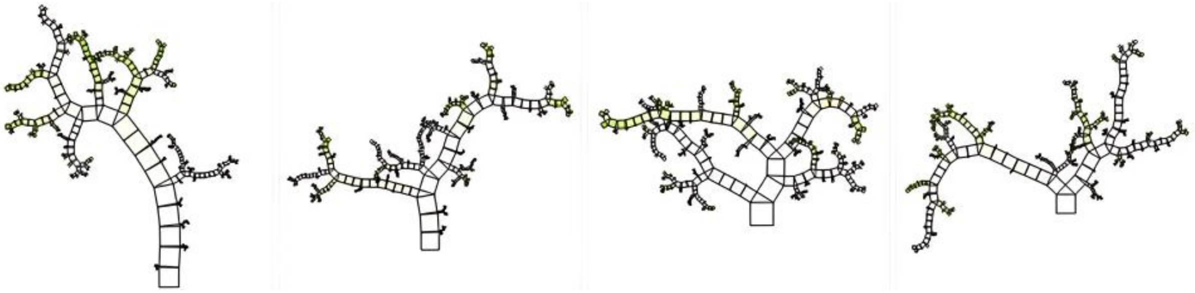


Figure 5. Example of Pythagorean random forest target class cumulative hotspot confidence level.

The model was designed to classify hotspots into three confidence level classes: low, medium, and high. Table 6 summarised the performance metrics of the classification model for each category. For the low-confidence category, the model achieved a high classification accuracy (CA) of 0.911; however, it exhibited a low F1 score (0.113) and recall (0.066), indicating difficulties in identifying instances belonging to this category. In contrast, the medium-confidence category demonstrated stronger performance, with an F1 score of 0.804, precision of 0.739, and recall of 0.882, reflecting the model's capacity to classify this category effectively. For the high-confidence category, the model achieved an Area Under the Curve (AUC) score of 0.795 and an F1 score of 0.499, suggesting moderate performance. On average, across all categories, the model recorded an AUC of 0.732, a CA of 0.703, and an F1 score of 0.669, indicating its general capability to classify hotspot confidence levels.

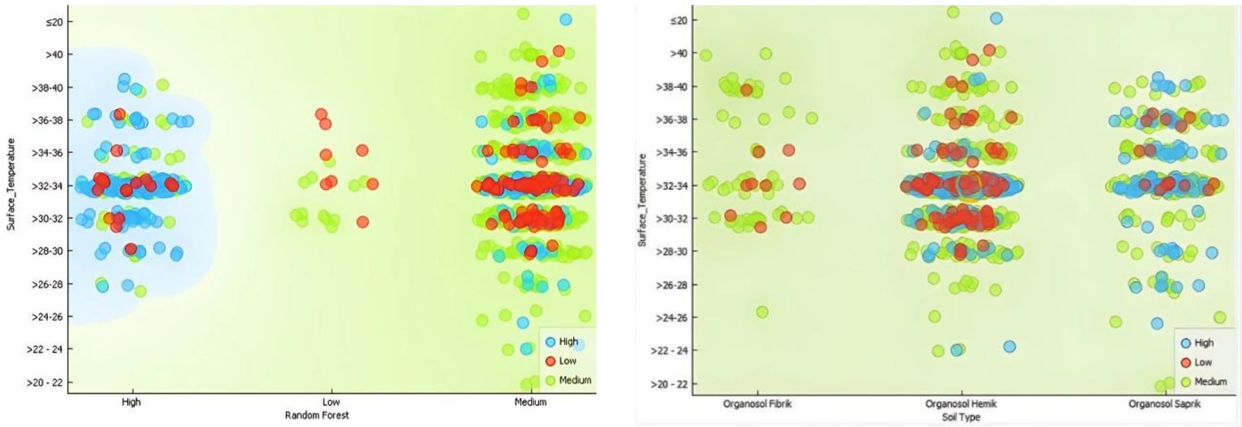
Table 6. Test score model random forest.

Target class	AUC	CA	F1	Precision	Recall
Low	0.651	0.911	0.113	0.400	0.066
Medium	0.707	0.715	0.804	0.739	0.882
High	0.795	0.779	0.499	0.574	0.442
Average over classes	0.732	0.703	0.669	0.669	0.703

An additional evaluation framework was used to assess the model's prediction performance, with results presented in Table 7. The low-confidence category achieved an impressive AUC score of 0.993, highlighting the model's ability to rank instances accurately. However, its F1 score (0.652) suggested balancing precision and recall challenges. The medium-confidence category exhibited excellent performance, with an F1 score of 0.925, precision of 0.887, and recall of 0.966. The high-confidence category achieved an AUC of 0.978 and an F1 score of 0.841, indicating strong performance but with room for improvement in balancing precision and recall. The strongest relationship between variables was observed between soil type and surface temperature, which were closely associated with hotspot confidence levels (Figure 6).

Table 7. Prediction performance score.

Target class	AUC	CA	F1	Precision	Recall
Low	0.993	0.887	0.652	1	0.484
Medium	0.973	0.887	0.925	0.887	0.966
High	0.978	0.887	0.841	0.867	0.816



(a) (b)

Figure 6. Relationships between variables surface temperature and hotspot confidence level (a); surface temperature and soil type (b).

Figure 6a illustrates the relationship between surface temperature and the classification results of the Random Forest model across three hotspot confidence levels (low, medium, high). Blue points (high confidence) are concentrated in the "High" category, predominantly associated with surface temperatures exceeding 30°C, indicating a strong link between high temperatures and high-confidence hotspots. Green points (medium confidence) are distributed in the "Medium" category, with surface temperatures ranging from 28°C to 32°C, while red points (low confidence) are found in the "Low" category, typically associated with surface temperatures below 28°C. This highlights the significance of surface temperature as a critical factor influencing hotspot confidence levels. Figure 6b depicts the relationship between surface temperature and soil type. Blue points (high confidence) are predominantly observed on organosol sapric soil types, with surface temperatures exceeding 30°C, while green points (medium confidence) are spread across all soil types, particularly organosol hemic, with temperatures ranging from 28°C to 32°C. Red points (low confidence) are primarily found on organosol fibric soil types, associated with surface temperatures below 28°C. This relationship emphasises that both soil type and surface temperature significantly affect hotspot confidence levels, with organosol sapric being more frequently associated with high-confidence hotspots.

Classification Model Evaluation

Table 8 shows the confusion matrix of the classification results for each category. The model successfully classified 57.4% of high-confidence instances, with a 17.5% misclassification to the medium-confidence category. The low-confidence category presented the most challenges, with only 40% of instances classified correctly, while the medium-confidence category achieved the highest accuracy at 73.9%. The class imbalance in the dataset was evident, with 944 instances of medium confidence compared to 353 high-confidence and 122 low-confidence instances.

Table 8. Confusion matrix.

		Predicted			
		High	Low	Medium	∑
Actual	High	57.4%	0.0%	17.5%	353
	Low	6.2%	40.0%	8.6 %	122
	Medium	36.4%	60.0%	73.9%	944
	∑	272	20	1127	1419

Discussion

This study successfully demonstrated that the RF algorithm can predict the confidence level of forest fire hotspots in peatlands with high accuracy, despite challenges in predicting low-confidence hotspots. One of the main findings of this research is its ability to handle large and complex datasets containing numerous interrelated variables, such as surface temperature, NDVI, and peat thickness. The RF model also effectively mitigates the overfitting issue often encountered with other methods such as DT, which aligns with the findings of Stojanova et al. [16], who noted that although RF performs well in terms of recall, its precision may be slightly lower compared to other methods. This study also underscores the relevance of satellite data, such as MODIS and Landsat 8, in forest fire analysis, which is further supported by Yu et al. [17], who emphasised the importance of satellite data in predicting forest fires in regions like Indonesia, which frequently experiences peatland fires.

However, as also observed in many previous studies, the model showed limitations in predicting low-confidence hotspots, which could be attributed to high variability in the data or unclear data quality. This issue is linked to the challenges faced by satellite-based models, which are affected by atmospheric conditions and limited data quality. This is consistent with the findings of Shofiana and Sitanggang [20], who reported that disturbances such as haze can lead to reduced confidence in identifying actual fires. Additionally, class imbalance in the dataset is a significant issue, with the medium-confidence hotspot category dominating, causing the model to classify data into this category more frequently, thus reducing the prediction accuracy for low- and high-confidence categories.

One of the key strengths of this study is the ability of the RF model to handle highly complex and large datasets, leading to more accurate predictions even with interrelated variables. The RF model proved effective in reducing overfitting, a common issue with models like DT, as found by Pang et al. [26], who demonstrated that RF could provide higher accuracy in predicting forest fires compared to other models such as Naïve Bayes or ID3. Another advantage of the RF model is its ability to manage data with various variables without requiring complicated preprocessing, making it an efficient choice in the context of big data analysis. However, a major weakness of this study is the class imbalance in the dataset, which causes the model to classify more data into the medium-confidence hotspot category. This aligns with the findings of Nurpratami and Sitanggang [21], who found difficulty in classifying certain categories in similar datasets using DT algorithms. To address this issue, the study by Rosadi et al. [27] used the SMOTE (Synthetic Minority Over-sampling Technique) method to balance the classes in the dataset, which improved the model's accuracy in predicting categories with fewer data.

Furthermore, the model is limited in predicting low-confidence hotspots, which reflects the challenges faced by satellite-based models, which are heavily influenced by atmospheric conditions and data quality limitations. When compared to other studies, one key difference is the success of Random Forest in handling larger and more complex data. For instance, the study by Dutta et al. [19] combined deep learning and ensemble methods to improve forest fire prediction and obtained better results compared to traditional methods such as KNN and DT. However, despite the challenges in predicting low- and high-confidence hotspots, the RF model in this study demonstrated superior performance compared to classical methods like Naïve Bayes or ID3. Research by Gigović et al. [28] and Unik et al. [29] also showed that RF outperforms Support Vector Machines (SVM) in mapping forest fire susceptibility, with better results in terms of prediction accuracy. On the other hand, while RF has proven effective in handling large data, the study by Ghali and Akhloufi [30] demonstrated that deep learning approaches could be more efficient in detecting and mapping forest fires using satellite data. This deep learning technique is more adaptive to dynamic changes in fires and provides higher accuracy in spatial analysis. Therefore, the use of deep learning methods, as outlined by Ghali and Akhloufi [30], could be the next step in optimising forest fire prediction in this study.

This research has significant implications for fire risk management, particularly in peatlands. The use of the Random Forest algorithm offers the ability to predict fires with a higher level of confidence, which can support more targeted and efficient fire mitigation measures. The results of this predictive model can help authorities plan and respond to fires more swiftly and effectively, focusing on hotspots with a greater potential for fire outbreaks. This is in line with the findings of Pang et al. [26], who emphasised the importance of using accurate prediction models to aid in more effective fire prevention planning. However, the challenges in predicting low-confidence hotspots suggest that the model needs further refinement, such as by improving data representation and introducing more comprehensive variables. The research by Ghali and Akhloufi [30], which integrates deep learning techniques, also offers the potential to enhance prediction accuracy, especially in detecting hotspots that are difficult to identify with traditional RF methods. Additionally, the research by Sirin and Medvedeva [31], which uses satellite data to detect peatland fires, provides insight into the potential use of multispectral data to improve the accuracy of fire mapping.

While this study offers valuable insights, several questions remain unanswered, presenting opportunities for further research. One of the major issues is the class imbalance in the dataset, which can affect the model's performance, particularly in predicting low-confidence hotspots. Rosadi et al. [27] shows that the SMOTE technique can improve model performance by balancing the classes in an imbalanced dataset, which could be applied to address this issue in future research. Furthermore, this study has not explored the possibility of integrating other temporal or spatial data that could enhance prediction accuracy, such as soil moisture information or more detailed weather data. Future research could consider using more complex models, such as deep learning or a combination of more advanced ensemble methods, to handle higher data complexity. Research could also explore the variables influencing hotspot confidence, such as the interaction between environmental factors and human activities, which could open up opportunities for developing more holistic and accurate models. Overall, while challenges in predicting low-confidence hotspots and issues related to class imbalance remain, this study shows great potential for the application of Random Forest in future forest fire mitigation efforts. Further research addressing these issues, as well as the exploration of additional variables, would be highly valuable in improving the accuracy of the model and supporting future forest fire mitigation initiatives.

Conclusions

The Random Forest algorithm was successfully applied to analyze the confidence level of hotspots in the peatlands of Riau Province, which are prone to forest and land fires. The results showed that this approach can predict fire potential with good accuracy, although there are challenges in predicting hotspots with low confidence levels. In particular, the model performed best in predicting medium-confidence hotspots, suggesting a clearer pattern or more representative data in this class. This can be attributed to the physical and environmental characteristics of the hotspots, such as surface temperature, NDVI, and peat thickness, which may be more consistent in medium-confidence hotspots. However, the main challenge is to predict low-confidence hotspots. This difficulty may be due to the high variability and lack of clarity in the data associated with low-confidence hotspots, which may require more variables or additional data to improve prediction accuracy. This suggests that machine learning models, such as Random Forest, must be accompanied by rich and diverse data to improve the prediction performance at all confidence levels. This research makes a significant contribution to the understanding and risk management of forest fires in peatlands. By improving the accuracy of hotspot predictions, fire mitigation efforts can be optimized, thereby reducing the environmental and socioeconomic impacts caused by forest fires. In addition, machine learning techniques, such as Random Forest, show great potential for analyzing complex environmental data, opening up opportunities for further research to develop more sophisticated and accurate models. Going forward, this study recommends collecting and integrating additional data, such as soil moisture, more detailed weather data, and information on human activities, to improve the quality and accuracy of the prediction model. Thus, this approach can be more effective in assisting the government and other stakeholders in making better decisions to prevent and suppress forest fires in peatlands.

Author Contributions

MU: Conceptualization, Investigation, Reviewing, Writing, and editing; **ISS**: Investigation, Methodology, and Writing an Original draft; **LS**: Research Design and Data Analysis; and **INS**: Conceptualization, Data curation, Writing–Reviewing and Editing, Funding Acquisition, and Project Administration.

Conflicts of Interest

There are no conflicts to declare.

Acknowledgment

The authors would like to thank the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia for providing financial support for this research.

References

1. Kumari, B.; Pandey, A.C. MODIS Based Forest Fire Hotspot Analysis and Its Relationship with Climatic Variables. *Spatial Information Research* **2020**, *28*, 87–99, doi:10.1007/s41324-019-00275-z.
2. Global Wetlands. The Global Wetlands Map Available online: https://www2.cifor.org/global-wetlands/ (accessed on 5 February 2022).
3. Osaki, M.; Nursyamsi, D.; Noor, M.; Wahyunto; Segah, H. Peatland in Indonesia. *Tropical peatland ecosystems* **2016**, 49–58.
4. Uda, S.K.; Hein, L.; Atmoko, D. Assessing the Health Impacts of Peatland Fires: A Case Study for Central Kalimantan, Indonesia. *Environmental Science and Pollution Research* **2019**, *26*, 31315–31327, doi:10.1007/s11356-019-06264-x.
5. Prayoto; Ishihara, M.I.; Firdaus, R.; Nakagoshi, N. Peatland Fires in Riau, Indonesia, in Relation to Land Cover Type, Land Management, Landholder, and Spatial Management. *J Environ Prot* **2017**, *8*, 1312–1332, doi:10.4236/jep.2017.811081.
6. KLHK Rekapitulasi Luas Kebakaran Hutan Dan Lahan (Ha) Per Provinsi Di Indonesia Tahun 2014-2019. Available online: http://sipongi.menlhk.go.id/hotspot/luas_kebakaran (accessed on 10 December 2019).
7. Panjaitan, A.; Pujiastuti, T.T.; Hanif; Ika, D.M. Satellite Perspective Of 2019 ’ S Forest Fire Disaster In Indonesia. Available online: http://www.bom.gov.au/research/aomsuc-10/presentations/S5-P9-PANJAITAN.pdf (accessed on 2 June 2023).
8. Zhang-Turpeinen, H.; Kivimäenpää, M.; Aaltonen, H.; Berninger, F.; Köster, E.; Köster, K.; Menyailo, O.; Prokushkin, A.; Pumpanen, J. Wildfire Effects on BVOC Emissions from Boreal Forest Floor on Permafrost Soil in Siberia. *Science of The Total Environment* **2020**, *711*, 1–12, doi:10.1016/j.scitotenv.2019.134851.
9. Syaufina, L. Research Status on the Relationship between Forest Fire and Biodiversity and Human Dimension in Indonesia. Presented at the National Seminar on Promoting Research on Forest and Land Fires Mitigation, Adaptation and Impact to Human and Biodiversity, Bogor, Indonesia, 2015.
10. Magadzire, N.; de Klerk, H.M.; Esler, K.J.; Slingsby, J.A. Fire and Life History Affect the Distribution of Plant Species in a Biodiversity Hotspot. *Divers Distrib* **2019**, *25*, 1012–1023.
11. Hayasaka, H.; Yamazaki, K.; Naito, D. Weather Conditions and Warm Air Masses during Active Fire-Periods in Boreal Forests. *Polar Sci* **2019**, *22*, 1–8, doi:10.1016/j.polar.2019.07.002.
12. Endrawati. *Analisis Data Titik Panas (Hotspot) dan Areal Kebakaran Hutan dan Lahan tahun 2016*; Purwanto, J., Eds.; Dirjen PKTL-KLHK: Jakarta, Indonesia, 2016; ISBN 978-602-61455-3-6.
13. Ranabhat, S.; Pokhrel, A.; Neupane, A.; Singh, B.; Gahatraj, S. Forest Fire Risk Assessment and Proposal for Fire Stations in Different Geographical Regions of Central Nepal. *Journal of Forest and Livelihood* **2022**, *21*, 46–59, doi:10.3126/jfl.v21i1.56585.
14. Nadeem, K.; Taylor, S.W.; Woolford, D.G.; Dean, C.B. Mesoscale Spatiotemporal Predictive Models of Daily Human- and Lightning-Caused Wildland Fire Occurrence in British Columbia. *Int J Wildland Fire* **2020**, *29*, 11–27, doi:10.1071/WF19058.
15. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012; ISBN 0262304325.
16. Stojanova, D.; Kobler, A.; Ogrinc, P.; Ženko, B.; Džeroski, S. Estimating the Risk of Fire Outbreaks in the Natural Environment. *Data Min Knowl Discov* **2012**, *24*, 411–442.
17. Yu, B.; Chen, F.; Li, B.; Wang, L.; Wu, M.; Mingquan, W.; Bo, Y.; Fang, C.; Bin, L. Fire Risk Prediction Using Remote Sensed Products: A Case of Cambodia. *Photogramm. Eng. Remote Sens* **2017**, *83*, 19–25.
18. de Angelis, A.; Ricotta, C.; Conedera, M.; Pezzatti, G.B. Modelling the Meteorological Forest Fire Niche in Heterogeneous Pyrologic Conditions. *PLoS One* **2015**, *10*, 1–17.
19. Dutta, R.; Das, A.; Aryal, J. Big Data Integration Shows Australian Bush-Fire Frequency Is Increasing Significantly. *R Soc Open Sci* **2016**, *3*, 1–11, doi:10.1098/rsos.150241.
20. Shofiana, D.A.; Sitanggang, I.S. Confidence Analysis of Hotspot as Peat Forest Fire Indicator. *J Phys Conf Ser* **2021**, *1751*, 1–10, doi:10.1088/1742-6596/1751/1/012035.

21. Nurpratami, I.D.; Sitanggang, I.S. Classification Rules for Hotspot Occurrences Using Spatial Entropy-Based Decision Tree Algorithm. *Procedia Environ Sci* **2015**, *24*, 120–126, doi:10.1016/j.proenv.2015.03.016.
22. Unik, M.; Sitanggang, I.S.; Syaufina, L.; Jaya, I.N.S. PM 2.5 Estimation Using Machine Learning Models and Satellite Data: A Literature Review. **2023**, *14*, 359–370.
23. BPS Provinsi Riau. *Provinsi Riau Dalam Angka 2020*; BPS Provinsi Riau: Pekanbaru, Indonesia, 2020.
24. Mogili, U.R.; Deepak, B.B.V.L. Review on Application of Drone Systems in Precision Agriculture. *Procedia Computer Science* **2018**, *133*, 502–509.
25. Choubin, B.; Soleimani, F.; Pirnia, A.; Sajedi-Hosseini, F.; Alilou, H.; Rahmati, O.; Melesse, A.M.; Singh, V.P.; Shahabi, H. Effects of drought on vegetative cover changes: Investigating spatiotemporal patterns. In *Extreme Hydrology and Climate Variability: Monitoring, Modelling, Adaptation and Mitigation*; Melesse, A.M., Abtew, W., Senay, G., Eds.; Elsevier: Philadelphia, PA, USA, 2019; Volume 2, pp. 213–222, ISBN 9780128159989.
26. Pang, Y.; Li, Y.; Feng, Z.; Feng, Z.; Zhao, Z.; Chen, S.; Zhang, H. Forest Fire Occurrence Prediction in China Based on Machine Learning Methods. *Remote Sens (Basel)* **2022**, *14*, 1–23, doi:10.3390/rs14215546.
27. Rosadi, D.; Arisanty, D.; Andriyani, W.; Peiris, S.; Agustina, D.; Dowe, D.; Fang, Z. Improving machine learning prediction of peatlands fire occurrence for unbalanced data using SMOTE approach. In Proceedings of the 2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA), Medan, Indonesia, 2021; pp. 160–163.
28. Gigović, L.; Pourghasemi, H.R.; Drobnjak, S.; Bai, S. Testing a New Ensemble Model Based on SVM and Random Forest in Forest Fire Susceptibility Assessment and Its Mapping in Serbia's Tara National Park. *Forests* **2019**, *10*, 1–21, doi:10.3390/f10050408.
29. Unik, M.; Sitanggang, I.S.; Syaufina, L.; Jaya, I.N.S. Stacking Regressor Model for PM2.5 Concentration Prediction Based on Spatiotemporal Data. *International Journal of Advanced Computer Science and Applications* **2025**, *16*, 1079–1086, doi:10.14569/IJACSA.2025.01601104.
30. Ghali, R.; Akhloufi, M.A. Deep Learning Approaches for Wildland Fires Using Satellite Remote Sensing Data: Detection, Mapping, and Prediction. *Fire* **2023**, *6*, 1–35, doi:10.3390/fire6050192.
31. Sirin, A.; Medvedeva, M. Remote Sensing Mapping of Peat-Fire-Burnt Areas: Identification among Other Wildfires. *Remote Sens* **2022**, *14*, 1–19, doi:10.3390/rs14010194.