Vol. 8, No. 2, April 2026

Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Logistic Regression Modeling of Peatland Fire Hotspots in Bengkalis District Using Integrated Environmental and Anthropogenic Drivers

Nur Hayati[1], Imas Sukaesih Sitanggang[2], Lilik Budi Prasetyo[3], Lailan Syaufina[4]

[1,2]School of Data Science, Mathematics and Informatics, IPB University, Bogor, 16680, Indonesia

[3]Dept. of Forest Resources Conservation and Ecotourism, Faculty of Forestry and Environment, IPB University, Bogor, 16680, Indonesia

[4]Dept. of Silviculture, Faculty of Forestry and Environment, IPB University, Bogor, 16680, Indonesia

Received:
January 6, 2026
Revised:
March 11, 2026
Accepted:
April 19, 2026
Published:
April 26, 2026

Corresponding Author:
Author Name*:
Nur Hayati
Email*:
nurh4y_nurhayati@apps.ipb.ac.id

DOI:
10.63158/journalisi.v8i2.1560

Abstract. Peatland fires occur almost annually in Bengkalis District, Riau Province, Indonesia, where peatlands cover about 65% of the area and contribute significantly to carbon emissions and regional haze, highlighting the need for improved fire risk prediction. This research aims to apply a probabilistic logistic regression approach to predict peatland fire hotspot occurrence and identify its key drivers. Hotspot data from 2015–2023 were derived from VIIRS satellite observations and classified into low (l), nominal (n), and high (h) confidence levels. Then hotspot confidence levels are classified into two scenarios: (1) the nh scenario (l = 0; n–h = 1) and (2) the h scenario (l–n = 0; h = 1), representing different fire thresholds. The predictor variable was modeled using anthropogenic and environmental, with multicollinearity testing to ensure model stability. The results show that the nh scenario performs better, with Nagelkerke R^2 = 0.0681, Hosmer–Lemeshow χ^2 = 5.7663, AUC = 0.69, and accuracy = 95.19%, indicating acceptable fit and moderate discrimination. Significant predictors include plantation land use, peat characteristics, and precipitation. These findings suggest that the approach can support peatland fire risk assessment, although further refinement is required.

Keywords: Bengkalis District, Fire Risk Assessment, Hotspot, Multicollinearity, Peatland Fires Prediction, Logistic Regression

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Indonesia is ranked fourth with the largest peatland in the world after Russia (137.5 million ha), Canada (113.4 million ha), and the United States (22.4 million ha) which is 14.9 ha [1]. Based on the total peatland area, Sumatra has 6.4 million ha, Kalimantan has 4.8 million ha, and Papua has 3.7 million ha [1]. Peatlands in Indonesia are able to store at least 57 billion tons of carbon, so that they have an important role in the world in relation to climate change mitigation [2], [3], [4]. A series of forest and peatland fires in Indonesia not only have an impact on the environment, but can also trigger problems in the economic and health sectors [5], [6]. The 2015 forest and land fire in Sumatra had severe transboundary impacts, causing health problems in more than 100,000 people across Indonesia, Malaysia, and Singapore [7]. One of the regions frequently affected by peatland fires is Bengkalis District, which serves as the case study area in this research. Every dry season, this area often experiences a spike in hotspots that cause haze, ecosystem damage, health problems, and economic losses [8], [9]. One of the main characteristics of peatland is its ability to store embers below the ground surface, making fires that occur difficult to detect and extinguish quickly [10], [11]. In this region, peatlands cover around 65% of the total land area, making it highly susceptible to fire risks and associated environmental consequences [12]. In addition, Bengkalis District is located on the coast of Riau Province which is adjacent to several neighboring countries such as Malaysia and Singapore, so that if a fire occurs it will cause cross-border air pollution problems [13]. In principle, the government already has a series of policies and control strategies to minimize the impact of forest and land fires. However, forest and land fires still occur almost every year in Bengkalis District. One of the causes is the implementation of forest and land fire prevention strategies that are still less accurate.

Based on previous research [12], [14], the existing conditions of the Land Use and Land Cover (LULC) [15] of peatlands in Bengkalis District are 11 types consist of plantations (42.2%), forests (44.9%), shrubs (11.2%), dry fields (0.7%), and fields (0.6%). While the other 6 categories consist of mangrove, rice fields, bare land, lakes, ponds, and fish ponds which only range from 0.01% to 0.19%. The oil palm plantation concession area in Bengkalis District reaches of 290,280.4 Ha and Industrial Plantation Forest reaches of 155,000.5 Ha which are located in Peat Ecosystem Protection Function (FLEG), which should not be used as concession areas because they will cause fire hazards because most of them are

peatland areas. The phenomenon of land concessions in FLEG indicates the emergence of various new open access areas that increasingly allow for deliberate burning of peatlands [16], [17]. In addition, the dense network of roads and canals that have been built in the peat ecosystem in Bengkalis District can increase the vulnerability to damage the peat ecosystem which, if not monitored and maintained intensively, will tend to increase and become vulnerable to fires.

The impact of fire conversion that occurred in Riau Province was greatly influenced by various factors such as environmental biophysical conditions, socio-economic conditions of the community and the implementation of laws and policies related to space allocation [18], [19]. The variables that are correlated with the occurrence of fires, namely the characteristics of peatlands, rivers, roads, LULC (forest, plantation, settlements, agriculture), permits (industrial plantation, concessions plantation), and forest areas (production forests, limited production forests, protected forests, and other use areas) [20], [21]. Based on the accuracy test, the spatial logistic regression model that was built can predict the probability of forest and land fires by 90.96% in Riau Province [22]. In addition, anthropogenic factors in cultivating agricultural or plantation using the slash and burn method are one of the factors that cause fires in Riau Province [23], [24], [25], [26]. Based on the analysis of the distribution of hotspots, a fairly high distribution of hotspots was found up to a distance of 20 km^2 from dry land agriculture [27], [28]. So, the closer the distance from dry land agriculture, the greater the risk of fire, considering that land clearing activities often use fire.

The previous research used logistic regression to produced three peatland fire prediction models, namely October (t-3), November (t-2) and December (t-1) with an accuracy level of 85.16%, 85.15% and 85.35% respectively where there is a positive relationship between the probability of peatland fires and the 4-month moving average of NDMI and rainfall. In this research, the spatio-temporal data used consisted of burnt areas (MODIS - MCD64a1) during the El Nino phase (2002, 2004, 2006, 2009 and 2015), physiographic (river network and canal density), peat characteristics, anthropogenic proxies (LULC, road density, river density, and channel density), and monthly rainfall [24].

This research also uses the logistic regression algorithm because of its reliable ability to model binary events, namely between fire (1) and not fire (0). However, unlike previous

studies that used the burn area variable, this research switched to using the hotspot variable obtained from the VIIRS (Visible Infrared Imaging Radiometer Suite) satellite sensor since 2015-2023 and classified into low (l), nominal (n), and high (h) confidence levels, where n and h generally indicate fire occurrence. Hotspot data collection in the 2015-2023 period was conducted to obtain a sufficiently long temporal representation to capture the dynamics of peatland fire occurrences. This time span encompasses annual variations in climate conditions, including extreme dry periods such as the 2015 El Niño phenomenon, as well as rainfall fluctuations in subsequent years, allowing for a more comprehensive analysis of fire occurrence patterns. VIIRS data was chosen because it has advantages in terms of higher spatial resolution (~ 375 meters) compared to the MODIS sensor (~ 1 kilometer), and has a high observation frequency so that it can detect fire events more accurately and in near real-time [29], [30]. The use of hotspots in this research is based on the consideration that burn areas are reactive and only reflect areas that have been burned, making them less suitable for supporting early warning systems and prevention efforts. On the other hand, hotspots function as early indicators of fire events and can be used proactively for prediction and mitigation planning. In addition, hotspot data can be obtained more quickly without requiring a complex validation process such as in burn areas, which are generally only available sometime after the fire event has finished.

Despite extensive researches on peatland fire prediction in Riau, existing approaches remain constrained by several limitations. Many models rely on burn area data, which are inherently reactive and less suitable for early warning applications, and often apply a single threshold to define fire occurrence, thereby overlooking uncertainty in hotspot confidence levels. In addition, previous researches tend to emphasize predictive accuracy without sufficiently integrating anthropogenic, physiographic, peat, and climatic drivers within a consistent binary spatial framework. Given that peatland fires in Bengkalis District occur recurrently and are strongly influenced by both environmental conditions and human activities, there is a need for a more analytically coherent approach that explicitly links hotspot detection, predictor representation, and model structure.

To address these gaps, this research applies a logistic regression–based framework using multiyear VIIRS hotspot data (2015–2023) and classified into two scenarios: (1) the nh scenario (l = 0; n–h = 1) and (2) the h scenario (l–n = 0; h = 1), representing different fire

thresholds. The objective is to evaluate how these alternative definitions influence model performance while identifying key drivers of peatland fire occurrence. By standardizing predictor variables into binary raster form and integrating them with hotspot-based response data, the research provides a consistent and interpretable modelling structure. Although the approach does not propose a new algorithm, it contributes empirical insight into the role of hotspot classification in improving model discrimination and supports a more flexible and operationally relevant basis for peatland fire risk assessment and early warning strategies in fire-prone regions.

2. METHODS

The research method begins with a literature review of peatland and forest fires, followed by data collection on forest and land fires. Figure 1 depicts the outline of the proposed method.

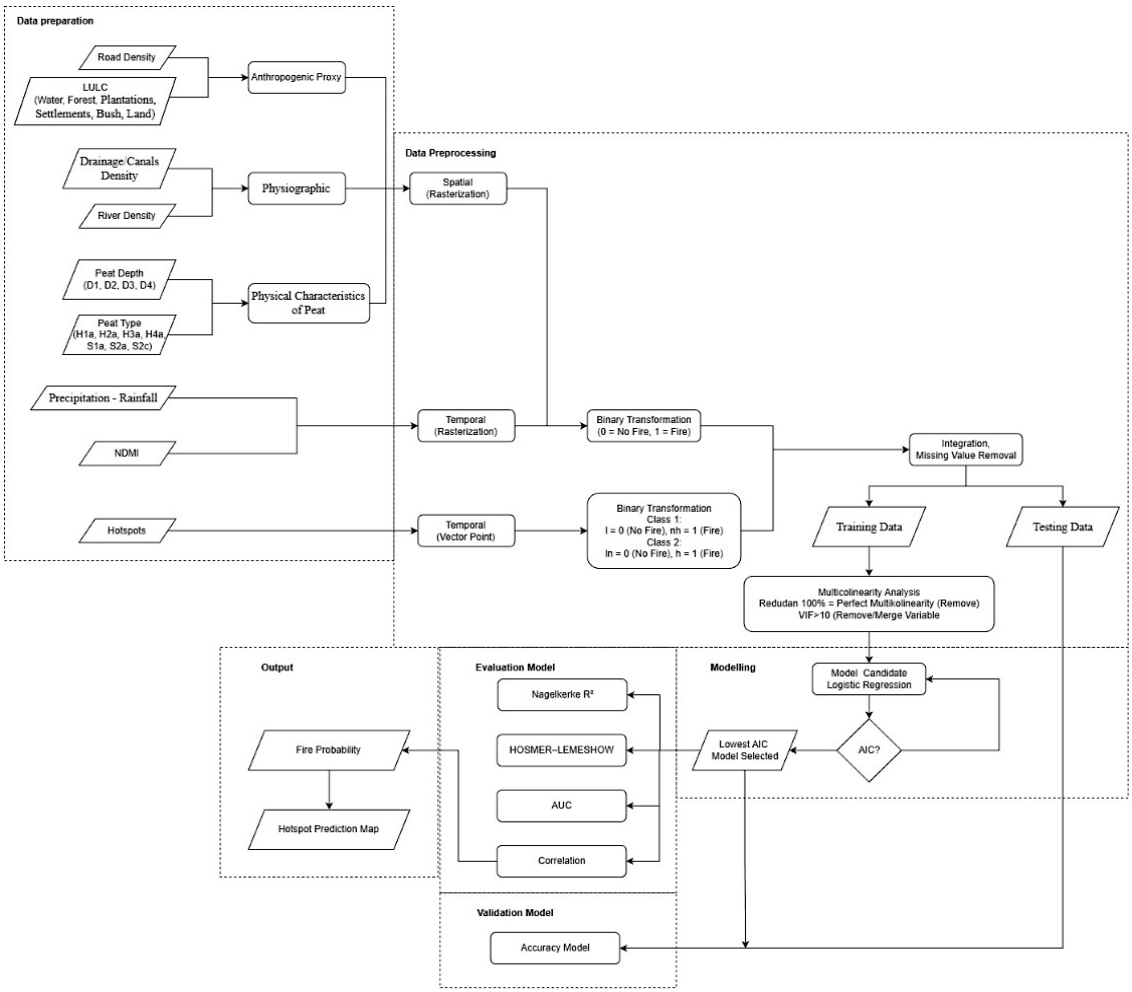

Figure 1. Research Methodology

The research methodology stages are as follows:

1) Data preparation, consisting of anthropogenic proxies including road density (vector line), land use and land cover (vector polygon) such as water, forest, plantations, settlements, bush, and land; physiographic data including drainage/canals density (vector line) and river density (vector polygon); peat physical characteristics (vector polygon) such as peat depth and type; precipitation - rainfall (raster), and NDMI (raster) as predictor variable and hotspots (vector point) as response variable.
2) Data preprocessing, all predictor variable were standardized into raster format to ensure spatial consistency in the modeling process. During the rasterization process with QGIS, vector data in the form of polygons and lines are automatically converted into raster grids with binary values (0 and 1), where the presence of a feature is assigned a value of 1 and its absence a value of 0. Meanwhile, for hotspot confidence level data consisting of l, n, and h, where n and h generally indicate fire occurrence, a classification process is carried out to produce binary data. There are two types of classification that can be carried out according to the identification of fire occurrence, namely (1) l = 0 and nh = 1, and (2) ln = 0 and h = 1. All predictor data were integrated with the temporal response data, missing values were removed, and the dataset was subsequently split into training and testing sets using an 80:20. Multicollinearity analysis was conducted to identify and remove or merge predictor variables exhibiting high correlation. Multicollinearity among predictor variables was evaluated using the Variance Inflation Factor (VIF). Variables with VIF values greater than 10 were considered to exhibit high multicollinearity and were excluded from further analysis.
3) Modelling, constructing a model using predictor variables refined through multicollinearity analysis with logistic regression algorithm. A comparison of Akaike Information Criterion (AIC) values was conducted to determine the best model (Lowest AIC).
4) Evaluation model, the optimal model was rigorously evaluated using the Hosmer–Lemeshow goodness-of-fit test, Nagelkerke pseudo-R^2 statistics, and the Area Under the Receiver Operating Characteristic Curve (AUC). Then, a correlation analysis was conducted to examine the relationships among variables. The Hosmer–Lemeshow goodness-of-fit test is applied to examine how well the predicted probabilities match the observed outcomes, indicating the overall model fit. The Nagelkerke R^2 is used to measure the explanatory power of the model, reflecting how much variation in

hotspot occurrence can be explained by the predictor variables. The Area Under the Receiver Operating Characteristic Curve (AUC) evaluates the model's ability to discriminate between fire and non-fire events. In addition, a correlation analysis is performed to further examine relationships among variables and support the interpretability of the model.

5) Validation model, is conducted to assess the reliability and predictive performance of the selected logistic regression model.
6) Output, the model produces a probability value (ranging from 0 to 1) for each spatial unit, representing the likelihood of hotspot occurrence under given environmental and anthropogenic conditions. These probability values are then translated into a Hotspot Prediction Map, which spatially visualizes the distribution of fire risk across the study area. The final output of the logistic regression analysis was a raster dataset that was subsequently reclassified according to defined probability thresholds. Areas with probability values between 0.3 - 0.8 were categorized as moderate, indicating a substantial potential for fire occurrence, while areas with probabilities > 0.8 were classified as high, reflecting a very high likelihood of peatland fire events.

2.1. Study Area

This research focuses on the peatland area within the administrative boundaries of Bengkalis District, Riau Province (Figure 2). Located a long Sumatra Island's eastern coastline, Bengkalis District covers approximately 7,773.93 km² [31]. Its boundaries include the Malacca Strait to the north, Siak District to the south, Meranti Islands and Karimun District to the east, and Dumai City along with Rokan Hilir and Rokan Hulu District to the west [32]. Bengkalis District lies on coastal lowlands only a few meters above sea level, consistent with regional Riau peatlands (base elevation ~6 ± 4 m a.s.l.) [33] Most of the soil in this District is organosol which contains a lot of organic material [34]. Rupat Island and Bengkalis Island are the two largest islands within the district, occupying 1,524.83 km² and 938.40 km² respectively [35]. The 11 Sub-Districts that make up Bengkalis District administratively are Bantan, Bandar Laksamana, Bathin Solapan, Bengkalis, Bukit Batu, Mandau, Pinggir, Rupat, Rupat Utara, Siak Kecil, and Talang Muandau [36]. Peatland areas in Bengkalis District are found in 8 Sub-Districts: Bantan, Bengkalis, Bukit Batu, Mandau, Pinggir, Rupat, North Rupat, and Siak Kecil [37].

Figure 2. Study Area of Research

The mapping of peatlands in Bengkalis District involved the overlay of data from Wetlands International Indonesia (WII) on peat depth and decomposition types, peat distribution maps from the Ministry of Agriculture, and the National Peat Ecosystem Function map, as defined by Ministerial Decree No. 130/2017. In general, as seen in Table 1, the results of the overlay analysis show that the area of peatland in Bengkalis District reaches 5,047.779 km² or 504,777.9 Ha, which covers 61.38% of the total administrative area of Bengkalis District.

Table 1. Comparison Between Administrative and Peatland Area in Bengkalis District

Sub-District	Administrative Area (Ha)	Peatland Area (Ha)	Percentage of Peatland to Administrative Area (%)
Bantan	31,596.03	27,990.81	88.59
Bengkalis	58,952.14	53,305.50	90.42
Bukitbatu	203,686.53	172,858.13	84.86
Mandau	116,656.62	37,159.71	31.85
Pinggir	194,702.85	76,239.67	39.16
Rupat	107,473.30	67,817.63	63.10
Rupat Utara	42,953.29	11,846.17	27.58
Siakkecil	66,421.73	57,560.29	86.66
Total	822,442.50	504,777.91	

2.2. Dataset

Various maps serve as the primary data sources in this research such as the administrative map of Bengkalis District (2018) was obtained from Badan Informasi Geospasial (BIG). Land Use and Land Cover (LULC) data for 2018 were sourced from BIG's Rupa Bumi Indonesia (RBI) maps and categorized according to the SNI 7645 classification system. Information on river, drainage, and road density was also extracted from the RBI maps published by BIG. The National Peat Ecosystem Functions map was based on the Decree of the Ministry of Environment and Forestry No. 130/2017, issued by the Directorate General for Forestry Planning and Environmental Governance. Additional peatland data were obtained from the Ministry of Agriculture (2017) and from Wetlands International Indonesia (WII), which provided detailed maps of peat depth and decomposition types. Furthermore, data on concessions and plantations for the year 2023 were derived from Global Forest Watch–World Resources Institute (GFW–WRI). Hotspot data from 2015 to 2023 were obtained from the FIRMS VIIRS database, rainfall data were sourced from CHIRPS, and NDMI data were obtained from the USGS Earth Explorer platform.

Figure 3 is a predictor variable that has been processed from the main data and converted into raster form consisting of drainage density, road density, river density, land cover (water, forest, soil, bush), land use (plantations, settlements), peat depth D1 (<50 cm), D2 (50-100 cm), D3 (100-200 cm), D4 (200-400 cm), peat type H1a (Hemic/Sapric with depth D2), H2a (Hemic/Sapric with depth D3), H3a (Hemic/Sapric with depth H4), H4a (Hemic/Sapric with depth >400 cm), S1a (Sapric/Hemic with depth D2), S2a (Sapric/Hemic with depth D3), S2c (Sapric/Mineral with depth D3), S3a (Sapric/Hemic with depth D4). Predictor variables that have been converted into raster form will have values of 0 and 1.

Meanwhile, Figure 4 and Table 2 show the distribution of hotspots with low (l), nominal (n) and high (h) confidence levels. Hotspot as an independent variable is used as an indicator of peatland fire. Fire detection with hotspots from remote sensing satellite data can be used as an early warning of a fire. Minister of Forestry Regulation P.12/Menhut-II/2009 defines a hotspot as a thermal anomaly used to indicate forest and land fires, where the detected location has a temperature higher than the surrounding environment. Hotspot identification is carried out using detection systems like VIIRS, which are

installed on the SUOMI NPP and NOAA-20 (JPSS-1) satellites [38], [39]. Compared to earlier systems like MODIS on the Aqua and Terra satellites, VIIRS offers higher resolution and can detect a greater number of hotspots. VIIRS provides data with higher resolution, which allows smaller fire detection and hotspots that may not be detected by other sensors [40].

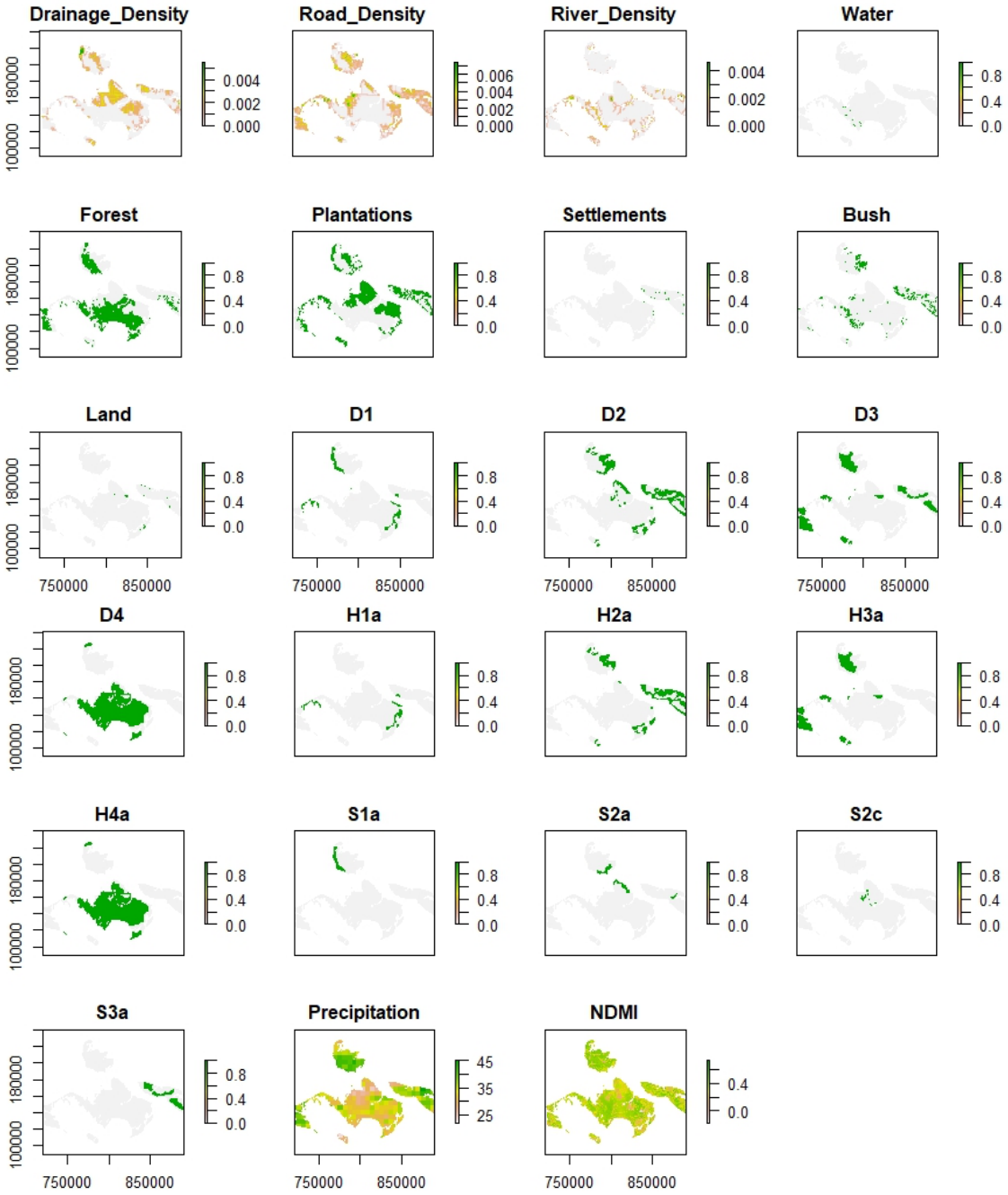

Figure 3. Predictor Variables

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

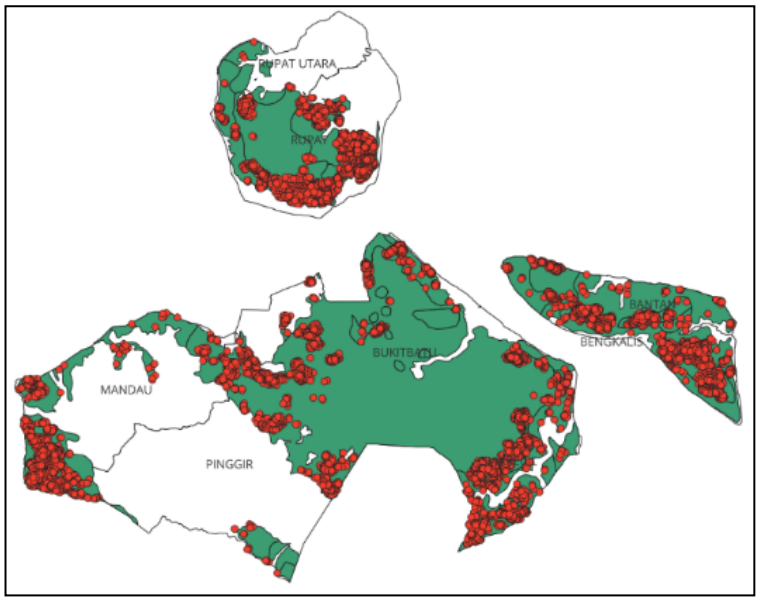

Figure 4. Distribution of Hotspots in Peatland Areas of Bengkalis District 2015-2023 with Confidence Level l-n-h

Table 2. Confidence Hotspots "l-n-h" Peatland Area of Bengkalis District in 2015 to 2023

Sub-District	Hotspots									Total
	2015	2016	2017	2018	2019	2020	2021	2022	2023	
Bantan	95	26		1	241	4	8	2	3	380
Bengkalis	901	449	4	24	845	74	370	1	36	2704
Bukitbatu	596	313		22	64	1	58		7	1061
Mandau	82	308	40	174	161	4	1		18	788
Pinggir	472	482	129	411	509	42	71	30	67	2213
Rupat	1706	403	1	124	1316	547	34	3	12	4146
Rupat Utara	9	3				7				19
Siakkecil	801	305	24	3	445	9		6	2	1595
Total	4662	2289	198	759	3581	688	542	42	145	12906

Table 2 show that the largest distribution of hotspots in the peatland areas of Bengkalis District occurred in 2015 and 2019. Meanwhile, the area with the highest average distribution of hotspots each year is in Bengkalis Sub-District. So, if a fire occurs in the peatlands of Bengkalis Sub-District, it will have a significant impact, considering that peatland areas are more flammable and difficult to extinguish, especially during the dry season.

Next, a conversion was carried out on 12906 hotspots, where the confidence level that allows a fire to occur is the one with the value n and h. So that the confidence level l is automatically valued at 0. While the confidence level n and h are valued at 1. However, in this research, it will be seen whether there is a significant difference if the value that is considered to have a high probability of a fire occurring is the one with the confidence level h. So that the hotspot data is grouped into two scenarios, namely: (1) l = 0 and nh = 1, and (2) ln = 0 and h = 1. After all datasets have been converted into binary form, the data from the predictor variables is combined with the response variables.

Next, the data for the response and predictor variables was merged. Missing data was then analyzed. If missing data was found, the missing data handling process was performed by deleting the data. This was done because the previous predictor data had different data types (before being converted to binary), making validation for missing data handling impossible.

2.3. Multicollinearity

After the data merging process, a multicollinearity test is performed. This analysis is performed to identify strong relationships or high correlations between predictor variables in a regression model. When two or more independent variables are highly correlated, the information provided becomes redundant, which can interfere with the parameter estimation process in the model. If the redundancy reaches 100% (perfect multicollinearity), the predictor variable must be removed because the coefficient cannot be estimated. Multicollinearity testing is essential before logistic regression modeling to maintain parameter stability and increase model validity. One commonly used method is the Variance Inflation Factor (VIF). A high VIF value (generally > 5 or > 10) indicates multicollinearity that needs to be addressed by combining highly correlated variables. The formula for multicollinearity is shown in Equation 1.

$$VIF = \frac{1}{1 - R_i^2} \tag{1}$$

Where R_i^2 = coefficient of determination of the regression results of X_i against all other predictor variables

2.4. Logistic regression

Logistic regression was employed to predict hotspot occurrence as a binary response variable based on a set of 23 predictor variables. This algorithm is well suited for modelling binary outcomes, as it estimates the probability of an event occurring in this case, the presence or absence of a hotspot given a combination of explanatory factors. By incorporating multiple environmental and land-related predictors, the logistic regression model enables an assessment of how these variables collectively influence hotspot occurrence and supports the development of a probabilistic prediction framework for peatland fire vulnerability (equation 2) [41]:

$$P = \frac{Exp(\beta_0 + \beta_1 X)}{1 + Exp(\beta_0 + \beta_1 X)} \quad (2)$$

where:

β_0 = intercept

β_1 = numerical coefficient of each predictor variable

X = predictor variable

Several logistic regression models were developed using different combinations of predictor variables, and their Akaike Information Criterion (AIC) values were subsequently compared to determine the optimal model. The AIC is defined as AIC = 2ln (L) + 2k, where L represents the maximum likelihood of the model and k denotes the number of estimated parameters. This criterion balances model goodness-of-fit and model complexity, thereby reducing the risk of overfitting and favouring more parsimonious models. The model with the lowest AIC value was selected as the best model and used for further analysis.

2.5. Evaluation and Validation Model

To ensure that the logistic regression model developed exhibits good performance and reliability, a comprehensive model evaluation process is conducted through verification and validation tests. Verification tests are performed to assess the goodness-of-fit and explanatory power of the model, including Nagelkerke R^2 indices, which describe how well the model explains the variability of the data, as well as the Hosmer–Lemeshow chi-square (χ^2) test to evaluate the agreement between observed and predicted values. Meanwhile, validation tests are employed to measure the predictive capability of the

model, using the Area Under the Receiver Operating Characteristic Curve (AUC) and overall accuracy, which indicate the model's ability to correctly classify events and non-events. The combination of these evaluation measures is essential to ensure that the model is not only statistically adequate but also robust and effective for practical application. The following are the formulas for these evaluation metrics [42]:

1) Nagelkerke R^2

$$R^2_{Nagelkerke} = \frac{R^2_{cox-snell}}{1 - exp\left(\frac{-2}{n}.lnL_0\right)} \quad (3)$$

where:

$R^2_{Cox-Snell} = 1 - \left(\frac{L_1}{L_0}\right)^{\frac{2}{n}}$

L_0 = likelihood null model

L_1 = likelihood full model

n = sample

Values range from 0 to 1, the higher Nagelkerke R^2 value, the better model's ability to explain data variation.

2) Homer-Lemeshow (χ^2) Test

$$\chi^2 = \sum_{g=1}^{G} \frac{\left(o_g - E_g\right)^2}{E_g} \quad (4)$$

where:

O_g= the actual number of observations in the gth group

E_g= predicted number of observations in the gth group

G = number of groups (usually 10)

3) Area Under the Receiver Operating Characteristic Curve (AUC)

$$AUC = \int_0^1 TPR(FPR)d\,(FPR) \quad (5)$$

where:

TPR = True Positive Rate

FPR = False Positive Rate

4) Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

where:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

Nagelkerke R^2 was employed as a pseudo-R^2 measure to assess the relative explanatory performance of the logistic regression model. Unlike the R^2 in linear regression, this metric does not represent the proportion of variance explained, as logistic regression does not involve continuous residuals or normally distributed errors and is estimated by maximizing likelihood rather than minimizing squared errors. Accordingly, a Nagelkerke R^2 value of approximately 0.10 or lower can still be considered reasonable in wildfire modeling, given that fire occurrence is a rare and inherently stochastic process influenced by complex and often unobserved factors. Model discrimination was further assessed using AUC, while calibration was evaluated using the Hosmer–Lemeshow test, ensuring that the predicted probabilities were consistent with observed hotspot occurrences.

3. RESULTS AND DISCUSSION

3.1. Integration and Binary Transformation Dataset

The prediction formula analysis process in this research was carried out using QGIS software version 3.32.1 and RStudio version 2026.01.1 which are open-source programming software that can support the big data computing process. As explained in the research methodology, the analysis process uses two classification scenarios to see a better hotspot prediction model. Prior to performing logistic regression analysis, it is necessary to ensure that the values of the response variable (hotspot) are in binary form. The binary transformation rules are defined as follows: in experiment 1, confidence level l is converted to 0, while n-h is converted to 1; in experiment 2, confidence levels l-n are converted to 0, while h is converted to 1. Subsequently, the response variable and predictor variables are integrated and transformed into raster data (binary form). In this research, experiment 1 is referred to as the nh confidence level scenario, while

experiment 2 is referred to as the h confidence level scenario. Table 3 presents the response and predictor variables used in the modeling process along with their corresponding value ranges.

Table 3. Value of Predictor and Response Variable

No.	Variable	Value	No.	Variable	Value
1	Hotspot	(0, 1)	13	Depth_D3	(0, 1)
2	Density_Drainase	(0 - 1)	14	Depth_D4	(0, 1)
3	Density_Road	(0 - 1)	15	Type_H1a	(0, 1)
4	Density_River	(0 - 1)	16	Type_H2a	(0, 1)
5	LC_Water	(0, 1)	17	Type_H3a	(0, 1)
6	LC_Forest	(0, 1)	18	Type_H4a	(0, 1)
7	LU_Plantation	(0, 1)	19	Type_S1a	(0, 1)
8	LU_Settlement	(0, 1)	20	Type_S2c	(0, 1)
9	LC_Bush	(0, 1)	21	Type_S2a	(0, 1)
10	LC_Land	(0, 1)	22	Type_S3a	(0, 1)
11	Depth_D1	(0, 1)	23	Precipitation	(23 - 45)
12	Depth_D2	(0, 1)	24	NDMI	(0 - 1)

The notation (0, 1) indicates that the variable is binary, where 0 represents the absence of a feature or event and 1 represents its presence. In this research, this format is used for both the response variable (hotspot: 0 = no fire occurrence, 1 = fire occurrence) and several predictor variables such as land use categories and peat characteristics that have been converted into discrete binary form. The notation (0–1) represents a continuous normalized range, where values are scaled between 0 and 1. This type of variable is typically derived from normalization or density-based measurements, such as road density, river density, drainage density, and NDMI. Values closer to 0 indicate lower intensity or influence, whereas values closer to 1 indicate higher intensity. In contrast, the notation (23–45) for precipitation indicates a continuous real-world value range, typically measured in millimetres. Unlike other variables that are binarized or normalized, precipitation is retained in its original continuous form to better capture variability in moisture conditions, which plays a critical role in influencing peatland fire occurrence.

Furthermore, Table 4 illustrates the distribution of hotspot data according to binary status and confidence level following the binary transformation process.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 4. Distribution of Hotspot on Binary Status and Confidence Level

Status	Confidence Level	
	h	nh
0	12393	650
1	513	12256
Total	12906	12906

In contrast, the h scenario exhibits a strongly skewed distribution toward non-fire observations (12,393 vs. 513), reflecting a more conservative classification that only considers high-confidence detections. While this reduces false positives, it also limits the model's ability to identify a wider range of fire occurrences, potentially reducing sensitivity and overall discriminative performance. This imbalance explains why the h model may achieve slightly higher accuracy, as predictions are dominated by the majority non-fire class, but performs worse in terms of AUC. Next, the integrated modeling dataset underwent a data cleaning process by removing observations containing missing values. As a result, the number of valid observations was reduced, and the final dataset used for further analysis is presented in Table 5.

Table 5. Train and Test Dataset

Item	Total
Train	9817 of 24 variabel
Test	2445 of 24 variabel
Total	12272 of 24 variabel

Prior to data partitioning, the integrated modelling dataset was subjected to a data cleaning process to remove observations containing missing values. Out of a total of 12,906 observations, 634 records were identified as having missing values and were subsequently excluded from further analysis. After this process, the final dataset consisted of 12,272 observations derived from 24 variables, which were then divided into a training set comprising 9,817 observations (80%) and a testing set comprising 2,445 observations (20%), as presented in Table 5. This data partitioning strategy ensures that the model is trained using the majority of the data while maintaining an independent subset for evaluating model performance and generalization capability.

3.2. Multicollinearity Analysis of Predictor Variable

Multicollinearity is a critical issue in regression-based modelling, as strong linear relationships among predictor variables can adversely affect parameter estimation, increase standard errors, and reduce model interpretability. Therefore, prior to constructing the logistic regression model, an assessment of multicollinearity among the 23 predictor variables was conducted to ensure model stability and reliable coefficient estimation. This analysis aims to identify predictors that exhibit high intercorrelation and to determine whether such variables should be removed or combined, thereby producing a more parsimonious set of predictors for subsequent modelling.

The results of the multicollinearity analysis indicate that six predictor variables, namely LC_Land, Depth_D4, Type_H4a, Type_S1a, Type_S2a, and Type_S3a, exhibit perfect multicollinearity, as these variables show exact linear dependence with other predictors and yield infinite or undefined VIF values. Such conditions prevent reliable parameter estimation and therefore require immediate treatment. Consequently, these six variables were excluded from subsequent modelling to avoid singularity problems and to ensure model stability. Based on the remaining predictor variables, a base logistic regression model was then formulated as follows:

$$\begin{aligned} model\_base \; = \; & status \sim Density\_Drainase \; + \; Density\_Road \; + \; Density\_River \\ & + \; LC\_Water \; + \; LC\_Forest \; + \; LU\_Plantation \; + \; LU\_Settlement \\ & + \; LC\_Bush \; + \; Depth\_D1 \; + \; Depth\_D2 \; + \; Depth\_D3 \; + \; Type\_H1a \\ & + \; Type\_H2a \; + \; Type\_H3a \; + \; Type\_S2c \; + \; Precipitation \; + \; NDMI \end{aligned}$$

Tables 6 and 7 present the VIF values for the nh and h scenarios, respectively, providing a basis for identifying highly correlated variables that may affect parameter estimation and model interpretability.

Table 6. VIF of Predictor Variable at Confidence Level nh

Variable	VIF	Variable	VIF
Density_Drainase	1.339164	Depth_D2	4.709761
Density_Road	1.160082	Depth_D3	2.885028
Density_River	1.095076	Type_H1a	1.000002
LC_Water	1.000003	Type_H2a	2.880476
LC_Forest	123.246234	Type_H3a	1.266383

Variable	VIF	Variable	VIF
LU_Plantation	105.313606	Type_S2c	1.000000
LU_Settlement	1.000000	Precipitation	2.161728
LC_Bush	74.663067	NDMI	1.060414
Depth_D1	1.490838		

Table 7. VIF of Predictor Variable at Confidence Level h

Variable	VIF	Variable	VIF
Density_Drainase	1.289741	Depth_D2	4.342163
Density_Road	1.192155	Depth_D3	2.975928
Density_River	1.145259	Type_H1a	1.515543
LU_Water	1.000003	Type_H2a	2.520777
LU_Forest	28.490537	Type_H3a	1.537939
LU_Plantation	22.995192	Type_S2c	1.000000
LU_Settlement	1.000000	Precipitation	1.868877
LU_Bush	16.943089	NDMI	1.071582
Depth_D1	1.815633		

Based on the results of the multicollinearity analysis (Table 6 and 7), there are eight logistic regression models were subsequently formulated as follows:

```
> m1 <- model_base
> m2 <- update(m1, . ~ . - LC_Forest)
> m3 <- update(m1, . ~ . - LU_Plantation)
> m4 <- update(m1, . ~ . - LC_Bush)
> m5 <- update(m1, . ~ . - LC_Forest - LU_Plantation)
> m6 <- update(m1, . ~ . - LC_Forest - LC_Bush)
> m7 <- update(m1, . ~ . - LU_Plantation - LC_Bush)
> m8 <- update(m1, . ~ . - LC_Forest - LU_Plantation - LC_Bush)
```

The multicollinearity analysis presented in Tables 6 and 7 reveals that several land-use variables, particularly LC_Forest, LU_Plantation, and LC_Bush, exhibit extremely high VIF values, indicating severe multicollinearity. This suggests that these variables are strongly correlated and may represent overlapping information related to land cover composition

and anthropogenic influence. Such redundancy can distort coefficient estimates, inflate standard errors, and reduce the interpretability of the logistic regression model. Notably, the magnitude of VIF values is substantially higher in the nh scenario compared to the h scenario, indicating that broader hotspot classification (nh) amplifies correlations among land-use variables.

To address this issue, eight alternative logistic regression models (m1–m8) were systematically constructed by selectively removing combinations of the highly collinear variables. This stepwise exclusion strategy is not merely a technical adjustment but reflects an effort to identify the most parsimonious model while preserving explanatory relevance. Models m2–m4 evaluate the individual contribution of each variable by removing one predictor at a time, whereas models m5–m8 assess the combined effects of removing multiple correlated variables. In particular, model m8 represents the most constrained specification, eliminating all three highly collinear predictors, thereby minimizing redundancy but potentially sacrificing important explanatory information.

3.3. Logistic regression analysis

Based on the comparison of Akaike Information Criterion (AIC) values among the eight logistic regression models, the best-performing model for the nh confidence level scenario is model 6, which yields the lowest AIC value of 3792.660, while for the h confidence level scenario, model 5 is identified as the optimal model with an AIC value of 3302.035. In model selection using AIC, a lower AIC value indicates a better trade-off between model goodness-of-fit and model complexity; therefore, the model with the minimum AIC is considered the most parsimonious and optimal. The AIC values for each candidate model are presented in Table 8.

Table 8. AIC Values of Logistic Regression Models for nh and h Confidence Level

Model	AIC	
	nh	h
m1	3795.560	3303.265
m2	3794.177	3304.000
m3	3794.742	3304.013
m4	3793.998	3304.948
m5	3796.256	3302.035

Model	AIC	
	nh	h
m6	3792.660	3304.009
m7	3797.906	3303.748
m8	3796.591	3302.130

These results further indicate that the optimal model structure differs between the two confidence scenarios, reflecting variations in how predictor variables contribute to fire occurrence under different classification thresholds. In the nh scenario, the selection of model 6—where LC_Forest and LC_Bush are excluded—suggests that these variables contain redundant or overlapping information with other predictors, particularly those related to land use and anthropogenic activities. Their removal improves model parsimony without substantially reducing explanatory power, indicating that fire occurrence in the nh scenario is better captured through a more simplified representation of land use dynamics.

In contrast, the h scenario identifies model 5 as optimal, which excludes LC_Forest and LU_Plantation, implying that high-confidence fire events are less dependent on broader land use categories and may be more strongly influenced by specific environmental or peat related conditions. This difference highlights that the relative importance of predictor variables is sensitive to the definition of fire occurrence, where the nh scenario captures a wider range of fire signals, while the h scenario reflects more selective and high certainty events. The results of the model evaluation are presented in Table 9.

Table 9. Evaluation of Models for nh and h Confidence Levels

	nh	h
pseudo-R^2		
Nagelkerke R^2	0.06810830	0.01716522
Hosmer and Lemeshow goodness of fit (GOF) test		
X-squared	5.7663	10.377
df	8	8
p-value	0.6734	0.2395
AUC	0.6869324	0.6007481
Accuracy	95.19%	95.64%

The comparative results clearly indicate that the nh classification scenario outperforms the h scenario in terms of overall model robustness and predictive relevance (Table 9). The nh model exhibits a substantially higher Nagelkerke R^2 (0.0681) compared to the h model (0.0172), indicating a greater explanatory capacity, which is particularly meaningful given the stochastic and rare-event nature of peatland fire occurrences. Although pseudo-R^2 values are inherently low in logistic regression models for rare events, the relative improvement observed in the nh scenario highlights its stronger ability to capture underlying fire dynamics.

In terms of model discrimination, the nh scenario achieves a higher AUC (0.6869) than the h scenario (0.6007), suggesting a more reliable capacity to differentiate between fire and non-fire events. Furthermore, both models pass the Hosmer–Lemeshow goodness-of-fit test ($p > 0.05$), indicating acceptable calibration. However, the lower X-squared value in the nh model (5.7663 vs. 10.377) reflects a closer alignment between observed and predicted probabilities.

While the h scenario yields a marginally higher classification accuracy (95.64%), this metric is less informative in imbalanced datasets, where the dominance of non-fire observations can inflate accuracy values. Therefore, greater emphasis is placed on AUC and pseudo-R^2 as more reliable indicators of model performance. The correlations among variables are presented in Table 10.

Table 10. Estimating Parameter of Logistic Regression Model for nh

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	5.35170	0.70182	7.625	2.43e-14 ***
Density_Drainase	-145.84482	79.49670	-1.835	0.066565 .
Density_Road	-31.43376	55.55140	-0.566	0.571496
Density_River	327.29133	283.97063	1.153	0.249094
LC_Water	13.32748	562.79392	0.024	0.981107
LU_Plantation	-0.27821	0.11290	-2.464	0.013736 *
LU_Settlement	12.64250	2399.54473	0.005	0.995796
Depth_D1	1.01955	0.35063	2.908	0.003640 **
Depth_D2	1.30906	0.24717	5.296	1.18e-07 ***
Depth_D3	-0.53553	0.15602	-3.432	0.000598 ***

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

	Estimate	Std. Error	z value	Pr(>|z|)
Type_H1a	12.75020	224.68564	0.057	0.954747
Type_H2a	-0.54682	0.22274	-2.455	0.014090 *
Type_H3a	1.60906	0.16134	9.973	< 2e-16 ***
Type_S2c	11.72614	1382.09468	0.008	0.993231
Precipitation	-0.07269	0.02026	-3.587	0.000334 ***
NDMI	0.23521	0.47088	0.500	0.617414

Table 10 show statistically significant effects on peatland fire occurrence. LU_Plantation, Depth_D1, Depth_D2, Depth_D3, Type_H2a, Type_H3a, and Precipitation are significant at the 95% confidence level, indicating their important roles in influencing fire occurrence. Depth_D1, Depth_D2, and Type_H3a have positive coefficients, suggesting that increases in these variables are associated with a higher probability of fire occurrence, whereas LU_Plantation, Depth_D3, Type_H2a, and Precipitation have negative coefficients, indicating a decreasing effect on fire probability. Other variables, such as Density_Road, Density_River, LC_Water, LU_Settlement, Type_H1a, Type_S2c, and NDMI, aren't statistically significant, implying limited influence within the model. Overall, the results demonstrate that peat depth, land-use characteristics, peat type, and precipitation are key factors affecting peatland fire.

Mathematically, the prediction formula for model nh can be written as follows:

$$P = \frac{Exp\begin{pmatrix} 5.35170 + (-145.84482) * Den\_Drain + (-31.43376) * Den\_Rd + 327.29133 * Den\_Rvr \\ +13.32748 * LC\_Water + (-0.27821) * LU\_Plant + 12.64250 * LU\_Sett \\ +1.01955 * Dept\_D1 + 1.30906 * Dept\_D2 + (-0.53553) * Dept\_D3 \\ +12.75020 * Type\_H1a + (-0.54682) * Type\_H2a + 1.60906 * Type\_H3a \\ +11.72614 * Type\_S2c + (-0.07269) * Precip + 0.23521 * NDMI \end{pmatrix}}{1 + Exp\begin{pmatrix} 5.35170 + (-145.84482) * Den\_Drain + (-31.43376) * Den\_Rd + 327.29133 * Den\_Rvr \\ +13.32748 * LC\_Water + (-0.27821) * L\square\square\_Plant + 12.64250 * LU\_Sett \\ +1.01955 * Dept\_D1 + 1.30906 * Dept\_D2 + (-0.53553) * Dept\_D3 \\ +12.75020 * Type\_H1a + (-0.54682) * Type\_H2a + 1.60906 * Type\_H3a \\ +11.72614 * Type\_S2c + (-0.07269) * Precip + 0.23521 * NDMI \end{pmatrix}}$$

The logistic regression equation for the nh model represents the probability of peatland fire occurrence as a function of multiple interacting predictors, where each coefficient reflects the direction and magnitude of influence on the log-odds of fire. Variables with positive coefficients, such as river density (Den_Rvr), LC_Water, LU_Plantation, LU_Settlement, peat types (H1a, H3a, S2c), and NDMI, indicate that an increase in these

factors is associated with a higher probability of fire occurrence. In particular, the strong positive coefficient of river density suggests that areas with more extensive hydrological networks may be more vulnerable to fire, potentially due to associated drainage and accessibility effects. Similarly, the positive contribution of plantation and settlement variables reinforces the dominant role of anthropogenic activities in triggering peatland fires.

Conversely, variables with negative coefficients, including drainage density (Den_Drain), road density (Den_Rd), peat depth classes (D1, D3), peat type H2a, and precipitation, indicate a decreasing effect on fire probability. The negative coefficient of precipitation is consistent with the physical role of moisture in suppressing ignition and fire spread, while the mixed effects of peat depth suggest that fire susceptibility varies across peat stratification, reflecting differences in fuel availability and subsurface burning potential. Notably, the presence of both positive and negative coefficients among peat-related variables highlights the non-linear and heterogeneous nature of peatland fire dynamics, where different peat conditions can either promote or inhibit fire occurrence.

After identifying the influential variables and establishing the logistic regression prediction model, a prediction map of peatland fire occurrence probability was generated as the main objective of this research. As shown in Figure 5, the resulting prediction map was classified into three probability levels [43]. Probabilities ranging from 30% to 80% (moderate) represent areas with a considerable potential for fire occurrence, whereas probabilities greater than 80% (high) indicate areas with a very high likelihood of peatland fire occurrence. The area extent for each probability class in each sub-district is presented in Table 11.

Areas with high fire probability (red) are predominantly concentrated in Bukit Batu, Siak Kecil, Bantan, and parts of Rupat, indicating localized hotspots of elevated vulnerability. These clusters suggest that fire occurrence is not randomly distributed but is strongly influenced by specific spatial configurations of anthropogenic pressure and peat characteristics, particularly in regions with intensive land-use activities and modified peat hydrology.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

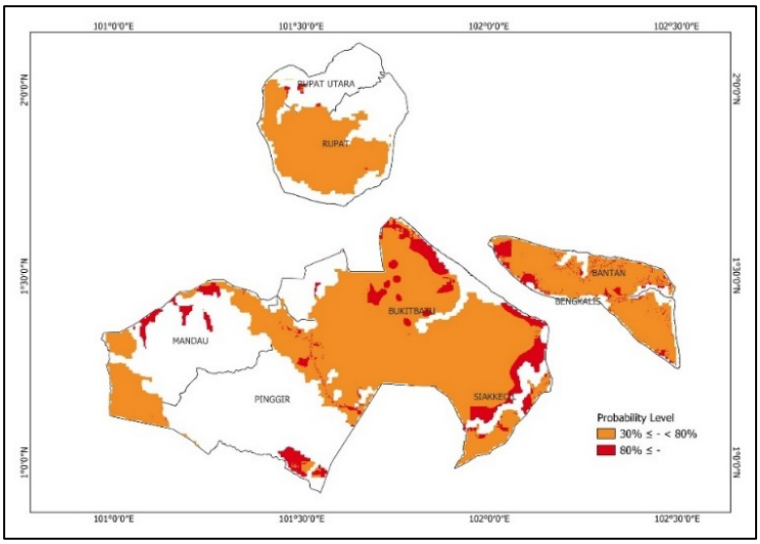

Figure 5. Probability of Peatland Fires Prediction in Bengkalis District

In contrast, moderate probability areas (orange) are more widely distributed across the region, covering substantial portions of Bengkalis, Pinggir, and Mandau, reflecting a broader landscape of latent fire susceptibility. This pattern indicates that while extreme fire risk is spatially concentrated, moderate-risk zones dominate the overall area, consistent with the earlier finding that most hotspots fall within the nominal–high (nh) classification. The widespread presence of moderate-risk areas highlights the importance of preventive and early detection strategies, as these zones may transition into high-risk conditions under favorable climatic or anthropogenic triggers.

Table 11. Area of Peatland Containing Hotspots with Fire Potential in Bengkalis District

Sub-District	Confidence Level		Total Area (Ha)
	Nominal (n) (Ha)	High(h) (Ha)	
Bantan	23,730.66	1,995.02	25,725.68
Bengkalis	43,865.79	5,676.31	49,542.10
Bukitbatu	149,715.72	17,304.82	167,020.53
Mandau	24,511.71	6,185.12	30,696.83
Pinggir	62,026.55	6,960.08	68,986.63
Rupat	63,588.67	50.53	63,639.20
Rupat Utara	7,264.73	655.67	7,920.40
Siakkecil	42,628.79	10,578.99	53,207.77

Sub-District	Confidence Level		Total Area (Ha)
	Nominal (n) (Ha)	High(h) (Ha)	
Total Area (Ha)	417,332.60	49,406.55	466,739.1519

The joint analysis of peatland distribution and hotspot occurrence reveals a non-linear and spatially differentiated relationship between peatland extent and fire activity in Bengkalis District (Table 1 and Table 11). While peatlands cover approximately 61.4% of the total administrative area (504,777.91 ha), the distribution of hotspots does not strictly follow peatland proportion alone, indicating that peat presence is a necessary but not sufficient condition for fire occurrence.

Sub-districts such as Bengkalis, Bantan, Siak Kecil, and Bukit Batu, which exhibit high peatland dominance (>84%), generally correspond to substantial hotspot areas, confirming that peat rich landscapes are inherently fire-prone. In particular, Bukit Batu stands out as the most critical hotspot cluster, with the largest peat extent (172,858.13 ha) and the highest hotspot area (167,020.53 ha), suggesting a strong coupling between peat availability and fire occurrence at large spatial scales.

However, this relationship is not uniform across all regions. For example, Siak Kecil, despite having a very high peat proportion (86.66%), exhibits a comparatively lower hotspot area (53,207.77 ha), indicating that additional controlling factors such as land use, accessibility, and human activities play a significant role. Conversely, Pinggir and Mandau, which have moderate peat coverage (39.16% and 31.85%, respectively), still show relatively high hotspot extents (68,986.63 ha and 30,696.83 ha), suggesting that anthropogenic pressures may amplify fire risk even in less peat dominated areas.

Furthermore, the dominance of nominal-confidence hotspots (~89.4%) over high-confidence hotspots (~10.6%) across all sub-districts indicates that fire risk is primarily distributed as moderate probability events across broad spatial extents, rather than being concentrated only in extreme hotspots. This reinforces the importance of early detection and preventive management strategies. Overall, the comparison demonstrates that peatland extent alone does not fully explain fire patterns, but must be interpreted in conjunction with anthropogenic and environmental drivers. These findings highlight the

need for integrated, location-specific fire management strategies that account for both peat characteristics and human-induced pressures, particularly in regions where moderate peat coverage coincides with high hotspot density.

3.4. Discussions

The findings of this research are consistent with previous wildfire and fire-occurrence studies using logistic regression, where low Nagelkerke pseudo-R^2 values are commonly reported. Studies in Spain and Finland show that, even with environmental, land-use, and anthropogenic predictors, logistic models often yield weak pseudo-R^2, sometimes below 0.01 that reflecting the stochastic and human-driven nature of fire ignition, particularly during transitional or low activity periods [44], [45]. These findings emphasize that pseudo-R^2 should be interpreted as a relative indicator of model improvement rather than explained variance. Accordingly, the Nagelkerke R^2 obtained in this research (0.0681) is comparable to, and in some cases higher than, values reported in earlier fire occurrence research and is therefore methodologically acceptable for hotspot-based peatland fire vulnerability modelling.

The influential variables identified in this study align closely with recent wildfire and hotspot modelling research, confirming that peatland fire vulnerability is driven by interactions between peat biophysical properties and human activities. In Bengkalis District, plantation land use, peat depth and type, and precipitation emerge as key predictors, indicating the combined effects of land management, peat susceptibility to drying, and moisture regulation. Similar patterns are reported in conservation-area studies, where land use, temperature, and proximity to roads dominate over topography [46]. Machine-learning-based studies in Indonesia and seasonal fire modelling in Thailand further highlight the importance of peat hydrology, climate, vegetation condition, and agricultural land use as joint determinants of fire likelihood [47], [48], [49]. Overall, these comparisons support an integrated framework combining land use, peat characteristics, climate, and anthropogenic pressure.

The methodological approach is well suited to peatland fire hotspot modelling, where fire occurrence is a binary and rare event. Logistic regression provides a transparent probabilistic framework, while the selection of predictors reflects established fire process knowledge. Importantly, multicollinearity analysis is essential to ensure

coefficient stability and interpretability given the strong correlations among environmental variables. Model evaluation therefore does not rely solely on Nagelkerke R^2 but emphasizes complementary metrics, including the Hosmer–Lemeshow goodness-of-fit test for calibration, AUC for discrimination, and classification accuracy for predictive performance. This multi-metric strategy ensures a robust and reliable assessment of model performance.

From a practical perspective, the model provides useful support for peatland fire management and early warning in Bengkalis District by identifying areas with higher fire vulnerability. The probabilistic outputs facilitate prioritization of monitoring, prevention, and mitigation efforts, particularly in plantation-dominated and hydrologically vulnerable peatlands. The integration of land-use, peat, and climatic variables, together with robust evaluation metrics, enhances operational reliability and confidence for real-world application. Consequently, the proposed approach can serve as an effective decision-support tool for local fire management agencies and may be adapted to other peatland regions with similar conditions.

4. CONCLUSION

This research demonstrates that logistic regression provides a methodologically robust yet inherently limited approach for predicting peatland fire occurrence in Bengkalis District under rare-event conditions. By transforming hotspot data into two confidence-based binary scenarios (nh and h) and addressing multicollinearity, the analysis shows that the nh model offers the best trade-off between discrimination and accuracy, outperforming the h model in terms of AUC while maintaining comparable accuracy. The nh model achieved Nagelkerke R^2 = 0.0681, Hosmer–Lemeshow p-value = 0.6734, AUC = 0.6869, and accuracy = 95.19%, indicating acceptable calibration and moderate discriminative ability. In contrast, the h model's slightly higher accuracy is less informative due to class imbalance and its lower AUC. Significant predictors include plantation land use, peat depth, peat type, and precipitation, with spatial predictions indicating that 466,739.15 ha (92.46%) of peatlands are potentially fire-prone. Nevertheless, the low pseudo-R^2 reflects the complex and stochastic nature of peatland fire dynamics, highlighting the need for further refinement through additional predictors, improved spatial–temporal resolution, and broader validation. Overall, the model provides a credible

baseline for probabilistic fire risk mapping and supports early warning and targeted mitigation strategies.

ACKNOWLEDGMENT

The authors gratefully acknowledge Universitas Nasional for financial support of this research. The authors also thank to Badan Informasi Geospasial (BIG), the Ministry of Environment and Forestry, the Ministry of Agriculture, Wetlands International Indonesia (WII), NASA FIRMS, CHIRPS, and USGS Earth Explorer for providing essential data used in this research. In addition, the authors sincerely appreciate the reviewers for their valuable comments and constructive suggestions that helped improve this manuscript.

REFERENCES

[1] S. Ritung, "Sosialisasi Peta Gambut BBSDLP 2019," in *Webinar: Perubahan Luasan Lahan Gambut dari Hasil Pemutakhiran Pemetaan Lahan Gambut*, Bogor, Indonesia, Dec. 2, 2020.

[2] M. A. Miller, P. Tonoto, and D. Taylor, "Sustainable Development of Carbon Sinks? Lessons From Three Types of Peatland Partnerships in Indonesia," *Sustainable Development*, vol. 30, no. 1, pp. 241–255, Feb. 2022, doi: 10.1002/sd.2241.

[3] L. Kiely, D. V. Spracklen, S. R. Arnold, E. Papargyropoulou, L. Conibear, C. Wiedinmyer, C. Knote 4, and H. A. Adrianto., "Assessing costs of Indonesian fires and the benefits of restoring peatland," *Nature Communications*, vol. 12, no. 1, Dec. 2021, doi: 10.1038/s41467-021-27353-x.

[4] C. S. Deshmukh, D. Julius, A. R. Desai, A. Asyhari, S. E. Page, N. Nardi, A. P. Susanto, Nurholis, M. Hendrizal1, S. Kurnianto, Y. Suardiwerianto, Y. W. Salam, F. Agus, D. Astiani, S. Sabiham, V. Gauci and C. D. Evans, "Conservation slows down emission increase from a tropical peatland in Indonesia," *Nature Geoscience*, vol. 14, no. 7, pp. 484–490, Jul. 2021, doi: 10.1038/s41561-021-00785-2.

[5] E. Quah, W. M. Chia, and T. S. Tan, "Economic impact of 2015 transboundary haze on Singapore," *Journal of Asian Economics*, vol. 75, Aug. 2021, doi: 10.1016/j.asieco.2021.101329.

[6] M. Taufik, M. Haikal, M. T. Widyastuti, C. Arif, and I. P. Santikayasa, "The Impact of Rewetting Peatland on Fire Hazard in Riau, Indonesia," *Sustainability* (Switzerland), vol. 15, no. 3, Feb. 2023, doi: 10.3390/su15032169.

[7] L. Hein, J. V. Spadaro, B. Ostro, M. Hammer, E. Sumarga, R. Salmayenti, R. Boer, H. Tata, D. Atmoko and J-P. Castañeda," The Health Impacts of Indonesian Peatland Fires," *Environmental Health*, vol. 21, no. 1, Dec. 2022, doi: 10.1186/s12940-022-00872-w.

[8] M. E. Harrison, N. J. Deerec, M. A. Imron, D. Nasire, Adul, H. A. Asti, J. A. Soler, N. C. Boyd, S. M. Cheyne, S. A. Collins, L. J. D'Arcy, W. M. Erb, H. Green, W. Healy, Hendri, B. Holly, P. R. Houlihan, S. J. Husson, Iwan, K. A. Jeffers , I. P. Kulu, K. Kusin, N. C. Marchant, H. C. Morrogh-Bernard, S. E. Page, A. Purwanto, B. R. Capilla, O. R. de R. Ortega, Santiano, K. L. Spencer, J. Sugardjito, J. Supriatna, S. A. Thornton, F. J. F. van Veen, Yulintine, and M. J. Struebig, "Impacts of Fire and Prospects for Recovery in A Tropical Peat Forest Ecosystem," *Proceedings of The National Academy of Sciences of The United States of America*, vol. 121, no. 17, Apr. 2024, doi: 10.1073/pnas.2307216121.

[9] M. A. Santoso, W. Cui, H. M. F. Amin, E. G. Christensen, Y. S. Nugroho, and G. Rein, "Laboratory study on the suppression of smoldering peat wildfires: effects of flow rate and wetting agent," *International Journal of Wildland Fire*, vol. 30, no. 5, pp. 378–390, May 2021, doi: 10.1071/WF20117

[10] G. Rein and X. Huang, "smoldering wildfires in peatlands, forests and the arctic: Challenges and perspectives," *Current Opinion in Environmental Science & Health*, Volume 24, 2021, 100296, ISSN 2468-5844, ht10.1016/j.coesh.2021.100296.

[11] M. J. Grosvenor, V. Ardiyani, M. J. Wooster, S. Gillott, D. C. Green, P. Lestari and W. Suri "Catastrophic impact of extreme 2019 Indonesian peatland fires on urban air quality and health," *Communications Earth and Environment*, vol. 5, no. 1, Dec. 2024, doi: 10.1038/s43247-024-01813-w.

[12] Fauziah, N. Hayati and L. B. Prasetyo, "Simulation of Land Use and Land Cover of Peatland Bengkalis Using QGIS." *International Journal on Informatics Visualization*, vol. 9, no. 1, Jan. 2025, doi: 10.62527/joiv.9.1.2432.

[13] H. D. Cahyani, M. M. Lestari, and L. Diana, "The Determination of State Baselines Post-Peat Abrasion on Bengkalis Island as Indonesia's Foremost Island in Terms of International Law of The Sea Perspective," *Multidisciplinary Indonesian Center Journal (MICJO)*, vol. 2, no. 2, pp. 1382–1401, Apr. 2025, doi: 10.62567/micjo.v2i2.639.

[14] S. I. Maulana, L. Syaufina, L. B. Prasetyo, and M. N. Aidi, "Pola Tutupan, Penggunaan, Serta Tantangan Kebijakan Perlindungan Ekosistem Gambut di Kabupaten Bengkalis," *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan (Journal of Natural Resources and Environmental Management)*, vol. 9, no. 3, pp. 549–565, Sep. 2019, doi: 10.29244/jpsl.9.3.549-565.

[15] R. Nedd, K. Light, M. Owens, N. James, E. Johnson, and A. Anandhi, "A synthesis of land use/land cover studies: Definitions, classification systems, meta-studies, challenges and knowledge gaps on a global landscape," Sep. 01, 2021, *MDPI*. doi: 10.3390/land10090994.

[16] R. Bambang Heryanto and Sukarman, "Peat Mapping on A Scale of 1:50,000 In Oil Palm Plantation Land, at The Peat Hydrological Unit in Musi Banyuasin Regency and The Implications of Its Use," in *BIO Web of Conferences, EDP Sciences*, Apr. 2024. doi: 10.1051/bioconf/20249905009.

[17] A. Y. Abdurrahim, A. H. Dharmawan, S. Adiwibowo, H. Yogaswara, and M. V. Noordwijk, "Actors, Access, Markets, and Values Involved in Oil Palm Expansion and Peatland Degradation in West Kalimantan, Indonesia," *Forest and Society*, vol. 9, no. 1, pp. 376–402, Jun. 2025, doi: 10.24259/fs.v9i1.34533.

[18] L. Juniyanti and R. O. P. Situmorang, "What causes deforestation and land cover change in Riau Province, Indonesia," *Forest Policy and Economics*, Volume 153, 2023, 102999, ISSN 1389-9341, doi: 10.1016/j.forpol.2023.102999.

[19] S. Mezbahuddin, T. Nikonovas, A. Spessa, R. F. Grant, M. A. Imron, S. H. Doerr and G. D. Clay, "Accuracy of tropical peat and non-peat fire forecasts enhanced by simulating hydrology," *Scientific Reports*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-022-27075-0.

[20] A. J. Horton, J. Lehtinen, and M. Kummu, "Targeted land management strategies could halve peatland fire occurrences in Central, Indonesia," *Communications Earth and Environment*, vol. 3, no. 1, Dec. 2022, doi: 10.1038/s43247-022-00534-2.

[21] B. S. Negara, R. Kurniawan, M. Z. A. Nazri, S. N. H. S. Abdullah, R. W. Saputra, and A. Ismanto, "Riau Forest Fire Prediction using Supervised Machine Learning," in *Journal of Physics: Conference Series*, 2020. doi: 10.1088/1742-6596/1566/1/012002.

[22] A. Yusuf, S. Siregar, and D. R. Nurrochmat, "Forest and land fires spatial model in Riau Province, Indonesia," 2018.

[23] A. Yusuf, Hapsoh, and S. H. Siregar, "Analisis Kebakaran Hutan dan Lahan di Provinsi Riau," *Dinamika Lingkungan Indonesia*, vol. 6, pp. 67–94, Jul. 2019.

[24] S. I. Maulana, L. Syaufina, L. B. Prasetyo, and M. N. Aidi, "Spatial Logistic Regression Models for Predicting Peatland Fire in Bengkalis Regency, Indonesia," *Journal of Sustainability Science and Management*, vol. 14, no. 3, pp. 55–66, 2019.

[25] M. Ohashi, A. Kameda, O. Kozan, M. Kawasaki, W. Iriana, K. Tonokura, D. Naito and K. Ueda, "Correlation of publication frequency of newspaper articles with environment and public health issues in fire-prone peatland regions of Riau in Sumatra, Indonesia," *Humanities and Social Sciences Communications*, vol. 8, no. 1, Dec. 2021, doi: 10.1057/s41599-021-00994-5.

[26] H. Hayasaka, "Fire Weather Conditions in Plantation Areas in Northern Sumatra, Indonesia," *Atmosphere (Basel)*, vol. 14, no. 10, Oct. 2023, doi: 10.3390/atmos14101480.

[27] K. A. Coskuner, "Assessing the performance of MODIS and VIIRS active fire products in the monitoring of wildfires: a case study in Turkey," *iForest Biogeoscience and Forestry*, vol. 15, no. 2, pp. 85–94, Apr. 2022, doi: 10.3832/ifor3754-015.

[28] Fauziah, L. B. Prasetyo, N. Saribanon, N. Hayati, "Vulnerability of Peatland Fires in Bengkalis Regency During the ENSO El Nino Phase Using A Machine Learning Approach," *MethodsX*, Volume 14, 2025, 103128, ISSN 2215-0161, doi: 10.1016/j.mex.2024.103128.

[29] C. I. Briones-Herrera, D. J. Vega-Nieva, N. A. Monjarás-Vega, J. Briseño-Reyes, P. M. López-Serrano, J. J. Corral-Rivas, E. Alvarado-Celestino, S. Arellano-Pérez, J. G. Álvarez-González, A. D. Ruiz-González, W. M. Jolly and S. A. Parks, "Near real-time automated early mapping of the perimeter of large forest fires from the aggregation of VIIRS and MODIS active fires in Mexico," *Remote Sensing (Basel)*, 2020.

[30] T. L. Schindler, "Active Fires as Observed by VIIRS, January–September 2021," Scientific Visualization Studio, NASA, 2021.

[31] L. Maulana, P. Suwarno, and T. Aris, "Global Warming and Its Impact on Mangrove Land Degradation on The North Coast of Bengkalis Island, Riau Province," *Journal of Agriculture (JoA)*, vol. 1 no. 2, pp. 2829–2421, Jul. 2022, doi: 10.47709/joa.v1n02.1574.

[32] Badan Pusat Statistik Kabupaten Bengkalis, *Kabupaten Bengkalis Dalam Angka 2024 (Bengkalis Regency in Figures, Vol. 14)*. Bengkalis, Indonesia: BPS Kabupaten Bengkalis, 2024.

[33] K. A. Hapsari, T. Jennerjahn, S. H. Nugroho, E. Yulianto, and H. Behling, "Sea level rise and climate change acting as interactive stressors on development and dynamics of tropical peatlands in coastal Sumatra and South Borneo since the Last Glacial Maximum," *Global Change Biology*, vol. 28, no. 10, pp. 3459–3479, May 2022, doi: 10.1111/gcb.16131.

[34] H. Junedi, A. K. Mastur, and AR Arsyad, "Study of the Critical Limits of the Ground Water for Peatland Fire Prevention," *Proceedings of the 3rd Green Development International Conference (GDIC 2020) Advances in Engineering Research*, vol. 205 hal 461-465, Atlantis Press, 2021.

[35] Bappeda Kabupaten Bengkalis, *Rencana Kerja Pemerintah Daerah (RKPD) Kabupaten Bengkalis Tahun 2023*, Bengkalis, Indonesia, Tech. Rep., *Peraturan Bupati Bengkalis* No. 36, 2022.

[36] Badan Pusat Statistik (BPS), "Luas Wilayah Provinsi Riau," *BPS Riau*, Agustus 11, 2023.

[37] Kementerian Lingkungan Hidup dan Kehutanan (KLHK), "Penetapan Peta Fungsi Ekosistem Gambut Nasional," *KLHK*, 2017.

[38] F. J. Vasconez, J. C. Anzieta, A. Vásconez Müller, B. Bernard, and P. Ramón, "A Near Real-Time and Free Tool for the Preliminary Mapping of Active Lava Flows during Volcanic Crises: The Case of Hotspot Subaerial Eruptions," *Remote Sensing (Basel)*, vol. 14, no. 14, Jul. 2022, doi: 10.3390/rs14143483.

[39] P. Sofan, F. Yulianto, and A. D. Sakti, "Characteristics of False-Positive Active Fires for Biomass Burning Monitoring in Indonesia from VIIRS Data and Local Geo-Features," *ISPRS International Journal of Geo-Information*, vol. 11, no. 12, Dec. 2022, doi: 10.3390/ijgi11120601.

[40] Y. Chen, , S. Hantson, N. Andela, S. R. Coffield, C. A. Graff, D. C. Morton, L. E. Ott, E. Foufoula-Georgiou, P. Smyth, M. L. Goulden and J. T. Randerson, "California wildfire spread derived using VIIRS satellite observations and an object-based tracking system," *Scientific Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1038/s41597-022-01343-0.

[41] S. Shabani, H. R. Pourghasemi, and T. Blaschke, "Forest stand susceptibility mapping during harvesting using logistic regression and boosted regression tree machine learning models," *Global Ecology and Conservation*, Volume 22, 2020, e00974, ISSN 2351-9894, 10.1016/j.gecco.2020.e00974.

[42] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Australia: *OTexts*, 2023.

[43] L. Giglio, W. Schroeder, J. V. Hall, and C. O. Justice, *MODIS Collection 6 and Collection 6.1 Active Fire Product User's Guide*, Version 1.0. College Park, MD, USA: University of Maryland, May 2021.

[44] S. Costafreda-Aumedes, Spatio-temporal analysis of human-caused fire occurrence patterns in Spain, Ph.D. dissertation, Univ. of Lleida, Lleida, Spain, 2017.

[45] H. Tanskanen and A. Venäläinen, "The relationship between fire activity and fire weather indices at different stages of the growing season in Finland," *Boreal Environment Research*, vol. 13, pp. 285–302, 2008.

[46] F. Kasim, A. Sabaruddin, and M. Hasan, "Forest fire susceptibility mapping using machine learning approaches in a conservation area," *Geosciences*, vol. 14, no. 1, pp. 1–20, 2024, doi: 10.2478/geosc-2024-0001.

[47] R. Suharyadi, H. Nugroho, and A. Nugraha, "Peatland fire susceptibility modeling using machine learning methods in Indonesia," in *Proc. Int. Symp. Remote Sensing and Intelligent Technology (ISRITI)*, Yogyakarta, Indonesia, 2020, pp. 1–6, doi: 10.1109/ISRITI51436.2020.9315359.

[48] Y. Zhang, Z. Liu, and X. Chen, "Machine learning–based wildfire susceptibility mapping using environmental and anthropogenic factors," *Remote Sensing*, vol. 17, no. 3, p. 3378, 2024.

[49] T. N. Phan, C. T. Kieu, and L. T. Nguyen, "Modeling seasonal fire probability in Thailand: A machine learning approach using multiyear remote sensing data," *Remote Sensing*, vol. 16, no. 1, pp. 1–25, 2024.