Article

Wildfire Hotspots Forecasting and Mapping for Environmental Monitoring Based on the Long Short-Term Memory Networks Deep Learning Algorithm

Evizal Abdul Kadir [1],* , Hsiang Tsung Kung [2], Amal Abdullah AlMansour [3], Hitoshi Irie [4], Sri Listia Rosa [1] and Shukor Sanim Mohd Fauzi [5]

[1] Department of Informatics Engineering, Universitas Islam Riau, Pekanbaru 28284, Indonesia
[2] Department of Computer Science, Harvard University, Cambridge, MA 02134, USA
[3] Department of Computer Science, King Abdulaziz University, Jeddah 22254, Saudi Arabia
[4] Center for Environmental Remote Sensing, Chiba University, Chiba 263-8522, Japan
[5] Faculty of Computer and Mathematical Science, University Teknologi MARA, Arau 02600, Malaysia
* Correspondence: evizal@eng.uir.ac.id

Abstract: Global warming is raising the earth's temperature, and resulting in increased forest fire events, especially in tropical regions with locations that are at high risk of wild and forest fires. Indonesia is a country in Southeast Asia that has experienced a severe number of wildfires, which have dangerous impacts on neighboring countries due to the emission of carbon and haze to the free air. The objective of this research is to map and plot the locations that consist of a significant number of fire hotspots and forecast the possible forest fire disasters in Indonesia based on the collected data of forest fires. The results of forecasting data are beneficial for the government and its policymakers to take preventive action and countermeasures regarding this wildfire issue. The Long Short-Term Memory (LSTM) algorithm, a deep learning method, was applied to analyze and then forecast the number of wildfire hotspots. The wildfire hotspot dataset from the year 2010 to 2022 is derived from the National Aeronautics and Space Administration's (NASA) Moderate Resolution Imaging Spectroradiometer (MODIS). The total number of collected observations is more than 700,000 wildfire data in Indonesia. The distribution of wildfire hotspots as shown in the results is concentrated mainly on two big islands, Kalimantan and Sumatra, Indonesia. The main issue is the peat type of land that is prone to spreading fire. Forecasting the number of hotspots for 2023 has achieved good results with an average error of 7%. Additionally, to prove that the proposed algorithm is working well, a simulation has been conducted using training data from 2018 to 2022 and testing data from 2021 to 2022. The forecasting result achieved a similar pattern of the number of fire hotspots compared to the available data in 2021 and 2022.

Keywords: wildfire hotspots; environment monitoring; LSTM algorithm; mapping and forecasting; Indonesia

1. Introduction

Global wildfires have become a critical environmental issue, with far-reaching consequences for ecosystems, human lives, and economies. While wildfires occur in various regions worldwide, research efforts have been undertaken to understand and address these challenges in specific areas. Several studies have investigated different aspects of wildfires and their management. The research of [1,2] focused on wildfire risk forecasting using the Weights of Evidence and Statistical Index models and explored forest fire risk forecasting through the application of case-based reasoning, while [3,4] proposed and developed a forest fire prediction model based on long short-term time-series networks and used deep learning algorithm approaches with unmanned aerial vehicle (UAV) images for forest fire smoke detection. Research conducted by [5] employed machine learning techniques

for flexible wildfire prediction. These studies provide valuable insights into wildfire dynamics and contribute to the development of effective strategies for fire management and prevention globally.

Wildfire is a major issue in tropical countries, particularly in Southeast Asian regions that are at high risk due to increasing temperatures. Additionally, the behavior of communities residing near wild and forested areas is one of the main contributing factors to fires, including deforestation and land clearing. Indonesia, located in the Southeast Asia region, experiences a tropical climate characterized by two seasons: rainy and dry. The landscape of Indonesia is predominantly covered by forests with peatland, especially on the Kalimantan and Sumatra Islands, where most of the land consists of peat, leading to a high risk of wildfires during the dry season. These disasters typically occur in the summertime when the dry soil conditions easily catch fire, particularly in proximity to hotspots. The behavior of residents in suburban or rural areas with dry land and small hotspots exacerbates the issue. Factors such as a lack of education, uncontrolled land clearing, and deforestation further contribute to the investigation of the disaster area. The impact of wildfires is particularly severe for communities near the fires, as the carbon emitted creates haze and air pollution, posing risks to human health, including respiratory illnesses. Furthermore, children and infants living in or near fire-prone areas may suffer from serious respiratory illnesses when the wind carries pollutants into the community. The effects of wildfires are not limited to humans; they also have detrimental impacts on animals and plants, as their habitats are destroyed by forest fires. Moreover, wildfires have long-term consequences such as global warming and climate change, as they strip the land of vegetation, leading to a reduction in oxygen generation from forests and wild trees. Figure 1 shows the wild and forest disaster in one of the states, Riau Province in Indonesia, and its impact on the habitat around the area as well as creating air pollution in neighbor states as well as neighboring countries. The research case is based in Riau Province Indonesia territory, with a latitude of 0.507319 and longitude of 101.444924.

Figure 1. Wildfire in Riau Province in Indonesia as area research case with latitude 0.507319 and longitude 101.444924.

Numerous relevant research studies have been conducted with various proposed techniques to alleviate or resolve the wildfire crisis as discussed in [6–8], where the wildfire method is fire data analysis and forecasting the hotspots using machine learning techniques. Additionally, the effect of climate change was influenced by the data analysis

and characteristics of the source data of the environment to be analyzed. The process of forecasting wildfire and the location of hotspot occurrence are related to the data from the meteorological agency as well as other factors of change in the environment. Another study based on the topic of forest and wildfires collected data from the data monitored environmental indicators that compare to the ground station normally detected or built by a special agency for disaster. The results of the analyses incorporate color codes to differentiate each indicator of wildfire. All these proposed methods to analyze and make predictions used simple applications as elaborated in [9,10]. The issue of wild and forest fire has been released for some of the equipment and the use of machine learning for prediction and location of the sources of hotspots, in addition to modeling and mapping of the scattering of fire. Observation and comparison according to weather data to determine the number of hotspots were simulated to achieve high-accuracy decisions in the scattering and number of fire hotspots in a location with a high possibility of wild and forest fires. Several groups have developed various models to observe the probability of potential forest fire occurrence considering both the natural and anthropogenic environmental properties. Recently, a new model has been proposed for investigating forest fire zones taking into account specific localization functions and evaluating critical fire parameters. Using this model and observations made by instrumentation on the remote sensing platform, a decision-making system has been proposed for the detection of areas prone to fires [11,12].

The method to analyze and identify the spread of the haze in the air due to wildfire has been elaborated and discussed by [13–15] to determine how much the area is being polluted by poor air quality. A deep learning algorithm called LSTM implements modeling to plot the pattern of the fire hotspot data, but the forecasting in this work only covers a small area or designated specific zone. The other work predicts and investigates fire datasets to using computerized analysis as one of methods to predict fire spread. An algorithm recurrent neural network (RNN) has the ability to analyze and integrate datasets for the propagation and prediction of fire hotspot spreading compared to others, while the use of RNN analyzes the dataset by the time series format. A technique to detect wild and forest fires in low levels used a ground sensing system by installing several sensors to detect the anomalies of environment indicators as discussed in [16–19]. Wireless sensor network (WSN) technology detects and monitors environmental change in ground sensing, such as temperature and humidity, due to global warming and dry seasons to obtain new datasets for analysis. Moreover, WSN technology collects accurate data in order to solve or overcome missing fire hotspot data, and the advantages of this technology, a direct detecting system compared to remote sensing, are useful for collecting direct fire data. Other results are mentioned in [20–22], which explore a method to predict the location of forest fire hotspots using a machine learning algorithm. Environmental monitoring is a method required to consider as an indicator to check air quality and the cleanness of the environment from pollution. Several studies [23–26] have examined how the environment can be monitored using multi-sensors to obtain detailed data from all the areas of forest and wilderness that are at high risk of fire. The use of WSN to collect data from the number of multi-sensors deployed in the forest is that each sensor has a coverage area to collect environmental data, reported to the system and alerted when a potential fire is detected. Remote monitoring from long distances or remote locations using a long-range wide area network (Loran WAN) is applied to have forests and wild locations monitored. Various sensors and different methods to detect forest fires are used because a location has a different type of forest, contour, land field, and type of soil; various models and techniques detect the potential of a fire event. Transmitting and reporting data to the backend system for further analysis and forecasting is obtained for the number of data collected.

The boreal and tropical forests undergo the most anthropogenic impacts. In particular, it has been noted that the wildfire phenomenon by lightning strike or by human actions is crucial for the forests' sustainability. New model simulation experiments showed that total burning of all coniferous forests up to 42° N resulted in an increase in atmospheric carbon by 21.7%, with subsequent global temperature increase by 4 °C. Wildfires are also

important for CH_4 emissions. In the modeling algorithms, the wildfires' CH_4 emissions are considered as 1–5 ($TgCH_4$/year) [27]. The issue of the atmospheric greenhouse effect has garnered significant interest, not only in scientific publications but also in mass media. On the one hand, there has been an undeniable overemphasis on the contribution of the greenhouse effect to global climate change. On the other hand, this heightened attention highlights the need to analyze the role of the greenhouse effect as a factor in climate change. Recent years have seen remarkable progress in analyzing observational data and the successful development of numerical climate modeling, providing a foundation for a fresh examination of the atmospheric greenhouse effect within the context of global climate change [28]. It has been observed that following the occurrence of forest fires, there is a decrease in solar ultraviolet radiation reaching the ground for several days. This decrease is attributed to the significant increase in concentrations of various air pollutants such as carbon monoxide, nitrogen oxides, ozone, and aerosol density during the forest fire event [29]. The application of remote sensing in environmental monitoring for sustainable development is discussed in the paper [30].

This research applied a method, namely the Long Short-Term Memory Networks (LSTM) deep learning algorithm, for several tasks on wildfires in Indonesia, such as:

- Mapping the wildfire hotspots in Indonesian territory based on the MODIS dataset.
- Plotting the distribution of wildfires and hotspots in the Indonesia region.
- Wild and forest fire forecasting to determine the potential hotspots in the future.
- Analyze the MODIS data on the area with high potential for fires.

The available Moderate Resolution Imaging Spectroradiometer (MODIS) dataset collected by NASA from 2010 to 2022 was used for this analysis to verify the model accuracy. The use of this dataset has limitations, such as MODIS data have a moderate resolution of 250 m and cannot detect small fires. In the analysis, the dataset is split into two categories, namely, a set of training data and a set of testing data. Data screening and filtering were implemented before running the analysis to obtain only the valuable dataset. The scope of this data analysis is limited to the Indonesian territory, since cases of forest fires frequently occur on two big islands, namely, Sumatra and Kalimantan. In this research, a new method was proposed for forecasting and mapping wildfire hotspots with the ability to achieve highly accurate hotspots compared to the previous works discussed, wherein a latest model of deep learning called LSTM contributes significant achievement to achieving precise locations of hotspots. Forecasting the number of hotspots for a future time in the specific zone, especially for Indonesian territory, is useful data for the respective agencies to propose further action. The Python programming language, which was used for this analysis and simulation, offers the advantage of being a high-level programming language capable of performing efficient calculations and accurately identifying wildfire hotspots based on the given dataset.

2. Material and Methodology

Natural disasters are a common phenomenon in the real world, according to region and geography, so forest and wildfire are disasters that commonly occur in most tropical countries with dense forests. Indonesia is one of the countries with a high risk of wildfire because of the land and area, with typical peatland and dry weather potentially catching fire. Indonesian government and community agencies supported by industry collaborate to counter and prevent this wildfire occurrence where possible, but some areas are not successful due to natural phenomena. Many studies have been conducted to find the main issues and root sources of fire, and to analyze the root causes of hotspots. This investigation uses a deep learning LSTM model for prediction and then plots prediction results to find and determine the scattering and distribution of the fire area especially in Indonesian territory. Earth data collected by NASA over the past 10 years are used in this analysis to plot and forecast the number of hotspots that occur in Indonesia. Figure 2 shows burn area data in Indonesia due to wild and forest fires in the years 2015 to 2021; 2015 is the

highest burn area, followed by the year 2019, and the remaining years average below 500,000 hectares.
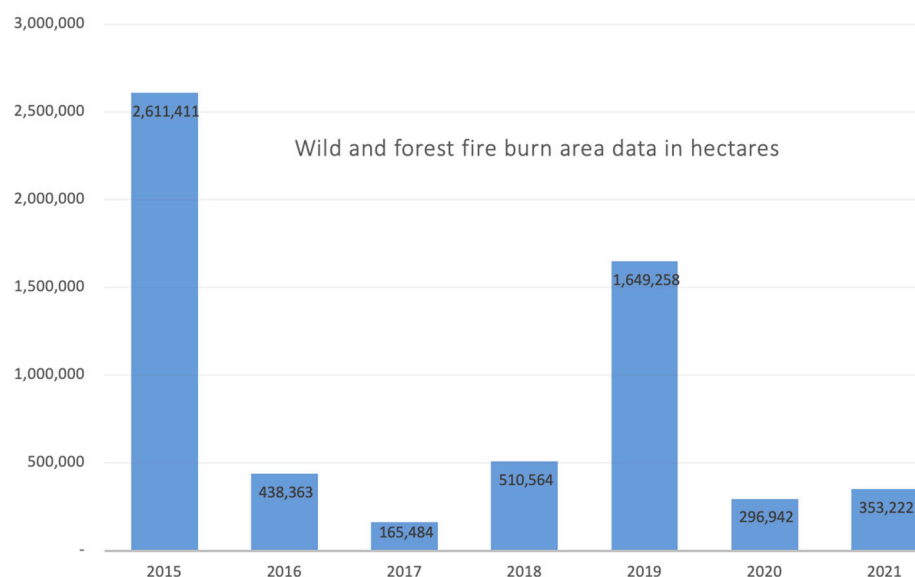
Figure 2. Wild and forest fire burn area data in Indonesia year 2015 to 2021 [30].

The data reveal that the cases of wild and forest fires in Indonesia primarily occur on Kalimantan and Sumatra islands due to the similar type of land, which is peatland, and the extensive forest coverage on both islands, posing a high risk of fires during the summer or dry season. The most significant case was observed in 2019, with numerous hotspots recorded in Indonesia, particularly in Kalimantan and Sumatra. However, in 2020, in response to the outbreak of the coronavirus disease (COVID-19), the government implemented mobility restrictions across the country, resulting in reduced activities, including transportation. Consequently, there was a notable improvement in air quality and a decrease in the number of wild and forest fire hotspots. Recent data from years after the COVID-19 pandemic demonstrate a significant decline in fire incidents and hotspots due to the restrictions on human mobility and limitations on outdoor activities. In this work, we will discuss the data from 2023 to compare the period before and after the COVID-19 pandemic and examine the impact on air quality resulting from government-enforced control measures and reduced pollution from transportation.

2.1. LSTM Algorithm

Currently, many techniques and algorithms are developed with good results and faster processing time. One of the popular deep learning algorithms is LSTM, which is the development of deep learning, a Recurrent Neural Network (RNN), initiated by Hochreiter and Schmidhuber [31]. The LSTM algorithm can analyze time series datasets to address the problem. This algorithm is also capable of learning long-term dependencies datasets and remembering the information for prolonged periods as a default. Figure 3 shows a basic architecture of an RNN-LSTM model, consisting of several main blocks called cells such as the input gate, output gate, and forget gate. The LSTM algorithm consists of two parts which, in the hidden state memory cell in the RNNs, the parts are C_t and the h_t work memory. Both memory cells are responsible for the retention of the sequence features. The sequence of the previous memory was controlled by the forgetting gate in the working memory, which is f, while h_t use is for the output, O is the control portion of the output gate, and C_t is current memory to be written as an input. The control portion of i for the state information of h_{t-1} and current input X_t is to write into the memory cell. All the types with three gates are not in a static condition, and the information in previous state as write h_{t-1} and current input of X_t are similarly determined as the nonlinear activation after

linear combination. LSTM has been successfully utilized in various applications, including brain sciences and the environment [3,32].
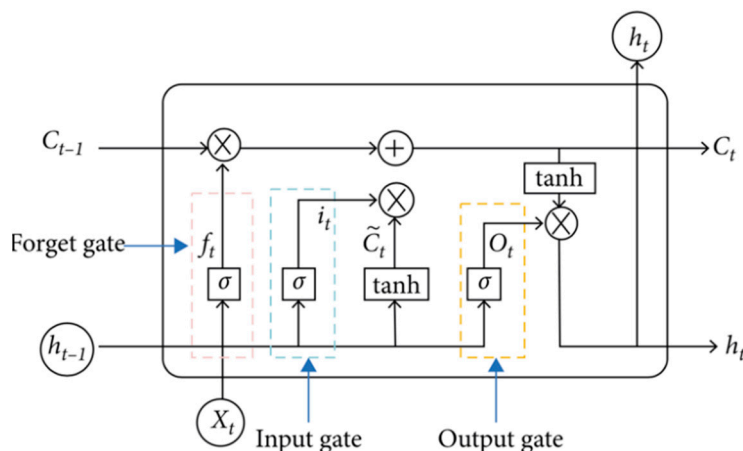

Figure 3. Structure of the RNN-LSTM algorithm.

The LSTM algorithm is a well-known deep learning technique that is capable of analyzing and addressing long-term dependencies in RNN by considering the historical data. On the other hand, other methods may have limitations in achieving accurate results when dealing with a long history of data or a large variety of data. Consequently, when there is a high volume and variety of data, along with a long history, these normal algorithms may not perform well in determining results. In such cases, the utilization of the LSTM algorithm, which effectively stores and utilizes information in long-term memory, provides more precise forecasting of the latest information. Implementing the LSTM algorithm as a fundamental method allows for the retention of information in the long-term stage, making it a commonly used technique for processing, analyzing, forecasting, and classifying basic time-series data [33].

The forecasting method to calculate the prediction number of wildfire hotspots used in the LSTM algorithm required further analysis and error justification to check the accuracy of the results. There are three blocks commonly explain in the LSTM structure which inside the block with dotted border which shows f_t, i_t, and O_t means for the forget, input and output gate. Many techniques can be used to check errors in forecasting analysis; for example, mean average error (*MAE*) to calculate average error, mean square error (*MSE*) to check errors in square, R^2 error analysis, a technique showing the proportion of the variant if forecasting number of dataset, and others as common error checking forecasting of dataset. These methods can be expressed as Equations (1)–(3) for *MAE*, *MSE*, and R^2, respectively:

$$MAE = \frac{\sum_1^N |yi - y_i'|}{N} \tag{1}$$

$$MSE = \frac{\sum_1^N (yi - y_i')^2}{N} \tag{2}$$

$$R^2 = 1 - \frac{\sum_1^N (yi - y_i')^2}{\sum_1^N (yi - y_{avg})^2} \tag{3}$$

where y_i is the number of fire hotspot datasets in actual time i, y_i' is the prediction number of hotspots in time i^{th}, and y_{avg} is the quantity of training dataset as the sample in analysis that check the number of errors; a metric regression model is used in this wildfire hotspot forecasting. All models may have different results from each other to calculate error percentage and error; hence, the model was applied to find the minimum error results in the forecasting of hotspots used as the best performance of dataset prediction. Error

calculation is very important in forecasting analysis to determine the performance level and accuracy prediction of an algorithm. If the error is too high above the threshold, then there is a need to check and calibrate the algorithm until the error percentage is within the acceptance level. Table 1 shows part of the dataset used in this forecasting with a full set, as, for example, the coordinate of the hotspot (latitude and longitude), date of incident, and total number of hotspots that occur.

2.2. Fire Dataset

The dataset of wildfire hotspots in this analysis was obtained from NASA's MODIS hotspot in Indonesia. Indicators in the dataset consist of 15 variables, and all the indicators are valuable for the analysis. The NASA Earth dataset is collected based on the image from the satellite to detect the active hotspot from the satellite station, because of the distance, whereas in some cases the quality of the image may have low quality and the hotspot is not counted in the number and is also affected by the total number of fire hotspots to be analyzed and forecasted. Then again, some sets of data that are missing due to complex parameters of earth data then strategize to overcome and complete the set of data by using the algorithm and predicting the missing database on training from the previous dataset. In this analysis and forecasting, only complete data achieve accurate prediction and determine the hotspot's location. Hotspot data in Indonesia—with a total of more than 11,000, as shown in Table 1—result in a highly accurate decision because of a high number of set data. The total data were collected in 10 years, from 2010 to 2022, and forecasted for the year 2023 [22]. Table 2 presents a set of NASA earth data after filtering, selecting only four parameters from 2010 to 2022 for analysis.

Refer to Tables 1 and 2 for the distribution of the fire hotspot dataset, which represents the daily occurrences of hotspots detected by satellites. On specific days, a higher number of hotspots are detected, requiring the data to be grouped and scaled into a single day to determine the total number of hotspots for each day. After the grouping process, the total number of datasets decreases significantly to only 4725 sets, compared to the previous count of up to 11,000 sets. These datasets represent a substantial amount of data collected over a span of 13 years, from 2010 to 2022. The purpose of this grouping and scaling is to optimize the forecasting and analysis of hotspot data. The data are presented in a line graph for informative analysis, showcasing the minimum and maximum occurrences of hotspots, as well as the average number over specific time intervals, allowing for conclusions to be drawn based on the plotted data. Moreover, to enable detailed analysis and forecasting, the data are further analyzed to provide monthly result values. Table 3 displays the optimized fire hotspot data, organized into three sets: the number of hotspot occurrences, the data itself, and the total number of hotspots in a single day. This optimization process enables a comprehensive assessment of the daily event data, allowing for meaningful comparisons on a monthly basis.

Table 1. Fire hotspot datasheet in Indonesia region from 2010 to 2022.

	Latitude	Longitude	Brightness	Scan	Track	Acq_Date	Acq_Time	Satellite	Instrument	Confidence	Version	Bright_T31	Frp	Daynight	Type
0	0.0211	116.87390	315.30	1.10	1.10	1 January 2023	251	Terra	MOOIS	42	6.2	295.60	8.70	D	0.0
1	0.48080	116.08060	312.30	1.00	1.00	1 January 2023	251	Terra	MOOIS	66	6.2	295.00	6.90	D	0.0
11	2.15090	117.49680	320.60	1.00	1.00	1 January 2023	550	Aqua	MOOIS	0	6.2	297.50	10.60	D	0.0
10	−8.10890	118.07430	319.30	1.00	1.00	1 January 2023	547	Aqua	MOOIS	0	6.2	300.80	9.10	D	0.0
8	−8.15960	117.58570	319.60	1.00	1.00	1 January 2023	547	Aqua	MOOIS	43	6.2	297.70	9.10	D	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
11,236	0.46716	108.99780	311.05	2.00	1.38	1 January 2023	319	Terra	MODIS	63	6.1NRT	295.90	15.66	D	NaN
11,235	−0.26051	109.49670	320.38	2.25	1.45	1 January 2023	319	Terra	MODIS	77	6.1NRT	294.89	41.98	D	NaN
11,250	−0.75817	101.51504	303.17	1.33	1.14	1 January 2023	1539	Terra	MODIS	54	6.1NRT	282.84	7.48	N	NaN
11,241	0.78876	110.72703	317.27	1.92	1.35	1 January 2023	557	Aqua	MODIS	73	6.1NRT	296.40	26.22	D	NaN
11,249	−0.75883	101.52123	301.76	1.33	1.14	1 January 2023	1539	Terra	MODIS	44	6.1NRT	283.03	6.28	N	NaN

Table 2. Fire hotspot used dataset in years 2010–2022.

No.	Latitude	Longitude	Date	Total
0	0.02110	116.87390	1 January 2010	42
1	0.48080	116.08060	1 January 2010	66
11	2.15090	117.49680	1 January 2010	0
10	−8.10890	118.07430	1 January 2010	0
8	−8.15960	117.58570	1 January 2010	43
.	.	.	.	.
11,236	0.46716	108.99780	31 December 2022	63
11,235	−0.26051	109.49670	31 December 2022	77
11,250	−0.75817	101.51504	31 December 2022	54
11,241	0.78876	110.72703	31 December 2022	73
11,249	−0.75883	101.52123	31 December 2022	44

Table 3. Daily number of hotspot data within years 2010–2022.

No.	Date	Total
0	1 January 2010	12
1	2 January 2010	12
2	3 January 2010	5
3	4 January 2010	14
4	5 January 2010	36
.	.	.
4721	25 December 2022	7
4722	27 December 2022	6
4723	29 December 2022	5
4724	30 December 2022	2
4725	31 December 2022	16

3. Fire Data Forecasting

The LSTM algorithm, a type of deep learning technique in computer programming, is highly regarded for its ability to analyze large volumes of data, particularly in the context of data analysis and time series forecasting. It falls under the RNN network type and is known for its efficient handling of long-term dependencies and memory retention. The LSTM model offers several advantages for data analysis, including its capacity to predict future events based on extensive historical data. With its four distinct layers and effective communication processing techniques, the LSTM algorithm organizes the model in a chain structure, enabling efficient data analysis. Figure 4 illustrates the block diagram of the data analysis scenarios used for forecasting wildfire hotspots in the future. The training and testing data are separated into individual blocks and undergo a rigorous process. Once the training data have demonstrated satisfactory performance, the testing data are utilized to forecast the number of hotspot occurrences in the future.

The first process in predicting and forecasting fire hotspot data is to build an LSTM algorithm model and then identify the input information. Unnecessary data in the filtering process can be ignored in the cell of the existing step. The identification process excludes the decoded data using the function of the sigmoid, taking the output model of the last part in LSTM unit h_{t-1} at the time $t-1$ and the current input X_t at time t. In addition, the function of sigmoid determines which parts of the previous old output should be removed. The forgetting gate is indicated by f_t, in which f is the vector value scale from 0 to 1, represents each number in cell state C_{t-1} as written in Equations (2)–(4).
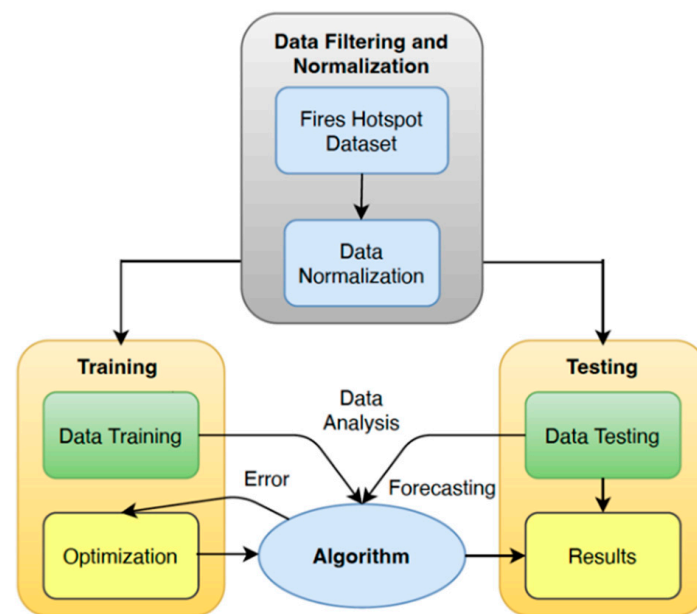
Figure 4. Block diagram analysis in data forecasting.

Referring to the collected dataset of more than 11,000 fire hotspots from the year 2010 to 2022, the first process is data filtering and normalization, deleting incomplete datasets or normalizing data to achieve a complete dataset. The grouping process resulted in total data to become 4213 datasets, as presented in Table 2, for the further machine learning process in data mapping and forecasting. The training data has a massive portion compared to the testing data, accounting for 80% of the total data used for training and then only 20% for testing data, as shown in Figure 2. The amount of training data is greater than the testing data in order to achieve high-accuracy decisions using the machine learning process. To evaluate the process and results obtained and then enhance the performance as well as the accuracy of forecasting data by minimizing the error in the results, an error calculation is used to check the percentage, to ensure that the error is below 10% as a normal indicator for data analysis error.

The LSTM algorithm is a type of recurrent neural network (RNN) architecture that is specifically designed to address the challenge of capturing and utilizing long-term dependencies in sequential data. Unlike traditional RNNs, which struggle with preserving and propagating information over long sequences, LSTM introduces specialized memory cells and gating mechanisms to selectively retain or discard information as needed. The key components of an LSTM network are the input gate, forget gate, memory cell, and output gate. These gates regulate the flow of information within the LSTM units. The input gate determines which portions of the input are important and should be stored in the memory cell. The forget gate decides which information from the previous time step should be discarded from the memory cell. The memory cell retains and updates information over time, allowing the network to remember long-term dependencies. Finally, the output gate determines the relevant information to be outputted from the memory cell.

The LSTM algorithm's ability to capture long-term dependencies makes it particularly well-suited for tasks involving sequential data, such as natural language processing, speech recognition, and time series analysis. It has demonstrated superior performance in various applications were preserving and utilizing information over long sequences is critical. By effectively handling the vanishing or exploding gradient problem, LSTM has revolutionized the field of deep learning and has become one of the most widely used techniques for modeling sequential data. Its architectural design allows for the efficient processing and analysis of long-term dependencies, making it a powerful tool in the realm of sequential data modeling and prediction. Figure 5 illustrates the neuron process of the LSTM model [34].
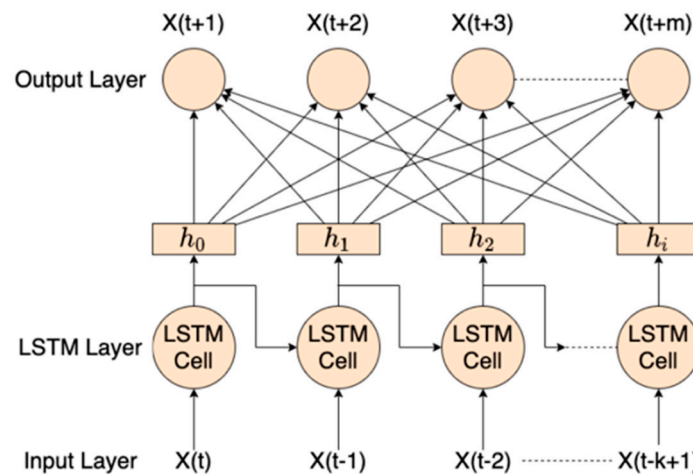
Figure 5. LSTM algorithm internal cell of the neuron model.

The final process calculates the error of the currently available dataset to forecasting results to check the percentage of error. Many techniques can be used to calculate the error. For example, Root Mean Square Error (RMSE) is a technique based on statistics normally used as a comparison between forecasting and real values of data. RMSE is often used to estimate how accurately the forecasting results match to history and reference values data on the relative scale of the dataset. Equation (4) elaborates that X_i and $X\prime_i$ are the actual fire hotspot dataset compared to the forecasting dataset at time t, X_i is the mean of real values of fire hotspot dataset, and N is the total amount of data. When the value of RMSE starts from a small number to zero, this implies that the LSTM algorithm produces reliable results.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - X\prime_i)^2} \tag{4}$$

On the other hand, the other method used to evaluate the forecasting results is mean absolute error (*MAE*), where M is the volume of observation dataset, X_i is the real values, and $X\prime_i$ is forecasting values. *MAE* is the standard of absolute errors as shown in Equation (5), which may well suggest the accuracy of the prediction error value.

$$MAE = \frac{\sum_{i=1}^{N}|X_i - X\prime_i|}{N} \tag{5}$$

Another method applied to calculate the error of forecasting data is the mean absolute percentage error (MAPE). MAPE represents the ratio of error compared to the real value in percent. Equation (6) depicts the formula to calculate MAPE error percentage in forecasting data of fire hotspots.

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|X_i - X\prime_i|}{p_a} \times 100\% \tag{6}$$

R-squared (R^2) is a technique used to verify whether a forecasting model is achieved and reflects the deviation of real dataset, normally in the range of [0, 1]. If R^2 is equal to 0, then the model fits poorly. If R^2 is equal to 1, then the model has no errors, as shown in Equation (7).

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(X_i - X\prime_i)^2}{\sum_{i=1}^{N}(X_i - \overline{X})^2} \tag{7}$$

Among all those methods used to evaluate prediction and forecasting errors, the RMSE and *MAE* evaluate based on the short term, i.e., hourly forecast. However, the MAPE method may have a calculation problem using a small denominator, as well as evaluating daily forecast basis. R^2 is one of the methods with detailed evaluation in a square calculation of all the datasets. The use of the LSTM algorithm for this process in

forecasting and mapping with a pseudo-code is shown in Algorithm 1. While *X* is the input dataset of hotspots from 2010 to 2022, *O* is the output data after forecasting results for the next future year; and in this scenario the future year is 2023 [35]. The process to obtain the results of the mapping and prediction of wild and forest fire specially in Indonesia region by computer simulation with analysis based on algorithm 1 as follow.

Algorithm 1: FORE-LSTM

Input: data stream $X = \{x_1, x_2, \ldots, x_i, \ldots\}, x_i \in R$, epoch *I*,
          number of iterations *K*, error parameters σ, cycle index *N*,
          Number of decompositions *m*, White noise data *W*.
Output: Forecasting Result *O*.
1: $X \leftarrow x_1, x_2, \ldots, x_i, \ldots, x_i \in R$ // Input data
2: $O \leftarrow \{ \}$ // Output data
3: **for** each $I \in N$ **do**
4:      $H_i^+ \leftarrow x + W_i^+; H_i^- \leftarrow x + W_i^-$
5:     **for** $j \in I$ **do**
6:         $H_{ij}^+ \leftarrow FORE\left(H_i^+\right); H_{ij}^- \leftarrow FORE\left(H_i^-\right)$
7:     **end for**
8: **end for**
9: **for** $j \in m$ **do**
10:   $O \leftarrow \sum\left(LSTM\left(IMF_j\right)\right)$
11: **end for**

---

4. Results and Discussion

The mapping and forecasting results of fire hotspots, as referred to NASA Earth data, started in the year 2010, complete with coordinates (latitude and longitude), date and time, the confidence level of the hotspot, and other indicators detected as previously shown in Table 1. In this research, analysis of only four indicators is used for further processing, namely, the acquisition date (acq date), the coordinates, and the confidence level. The spreading of hotspot levels is classified into five categories or levels of confidence. This method is implemented to verify which fire hotspot has a high-risk occurrence of forest fires and which among them is at low risk without much impact. The five levels of classification from the top level are as follows: 81–100, indicated as red dots; 61–80, indicated as orange dots; 41–60, indicated as yellow dots; 21–40, indicated as green dots; and 0–20, indicated as blue dots. Figure 6 represents the results of mapping the distribution of fire hotspots in Indonesia for the year 2022 and year 2010 to 2021 as in Appendix A. Overall, the hotspots are more concentrated in Sumatra and Kalimantan. As a result, these islands present a higher risk of forest fires because of their geography and land type. Ultimately, the best strategy is to conduct a prediction based on the current history data along with forecasting to alert the government, respective institutions, and communities. The dotted shows in the map consist of 5 types which from green color for the lower risk of fire to the blue, yellow, pink and red color indicate the highest potential of forest fire risk with percentage from 80 to 100.

The total number of hotspots detected is categorized into five levels for specific evaluation; all the fire hotspots refer to earth data in Indonesia that have counted thousands of hotspots in accumulated time over a 10-year period. In this work, all the hotspot data are plotted onto a graph annually for a detailed presentation indicated by level, starting from January until December every year. Figure 7 shows the monthly distribution of hotspots for every level for the year 2022 and the years 2010 to 2021 as in Appendix B, where the black line represents the total number of hotspots. In general, the pattern and distribution of hotspot incidents are similar for most of the year and increase at the end of the year from September to December due to summer or the dry season, while earlier on, the number of hotspots looks normal. The maximum number of hotspots detected is between 600 and 700 for the peak season at the end of the year. The study is in the location of Indonesian

territory with borders in latitude 5.980133, 94.964324 and −10.537824, 94.964324 on the west and in the east with 5.980133, 141.105341 and −10.537824, 141.105341.
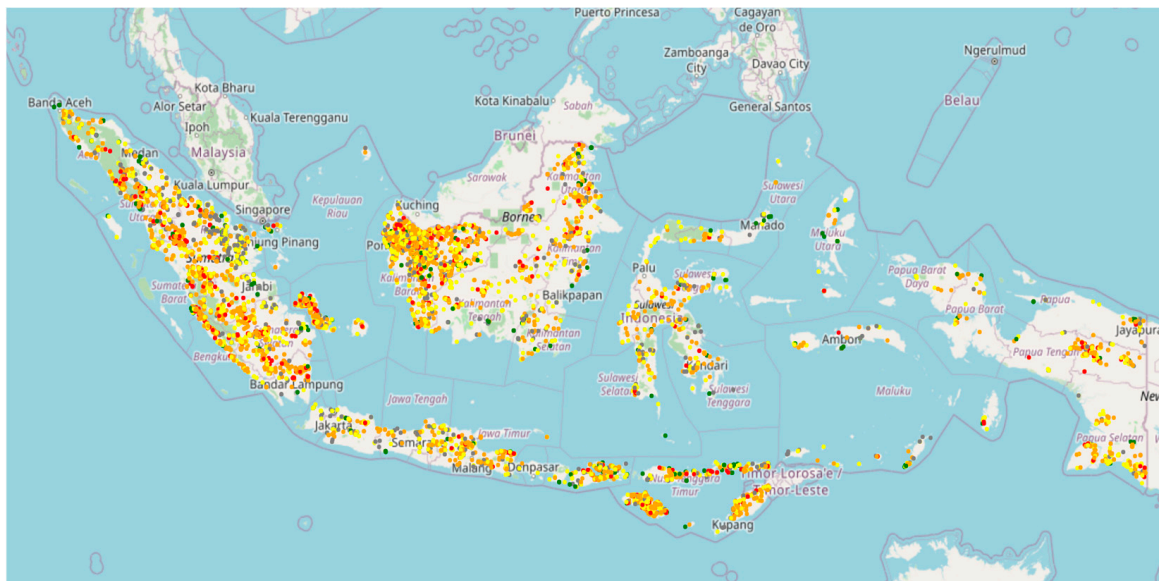
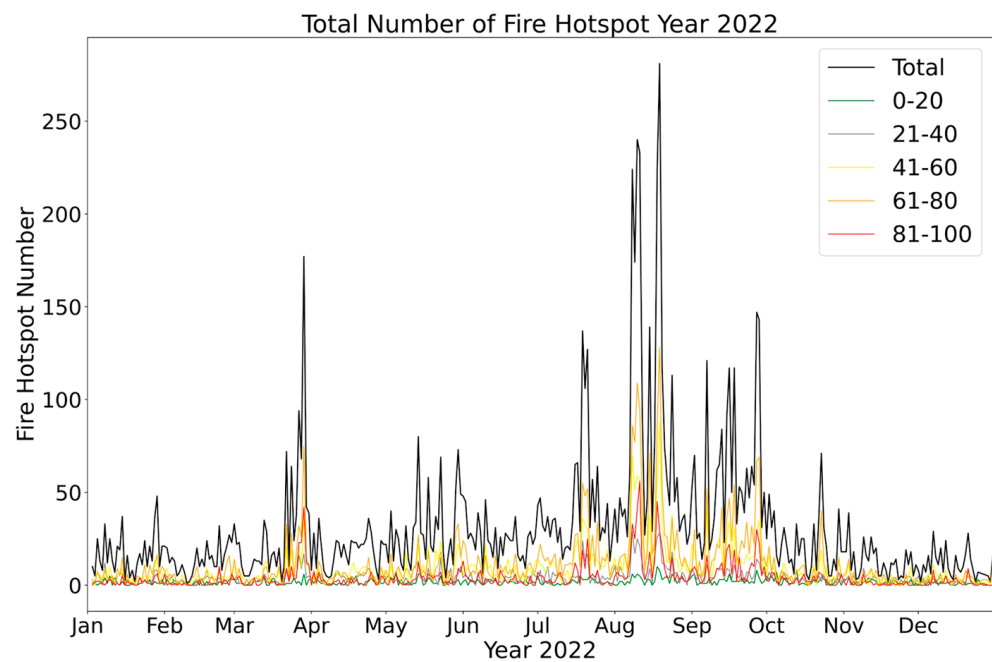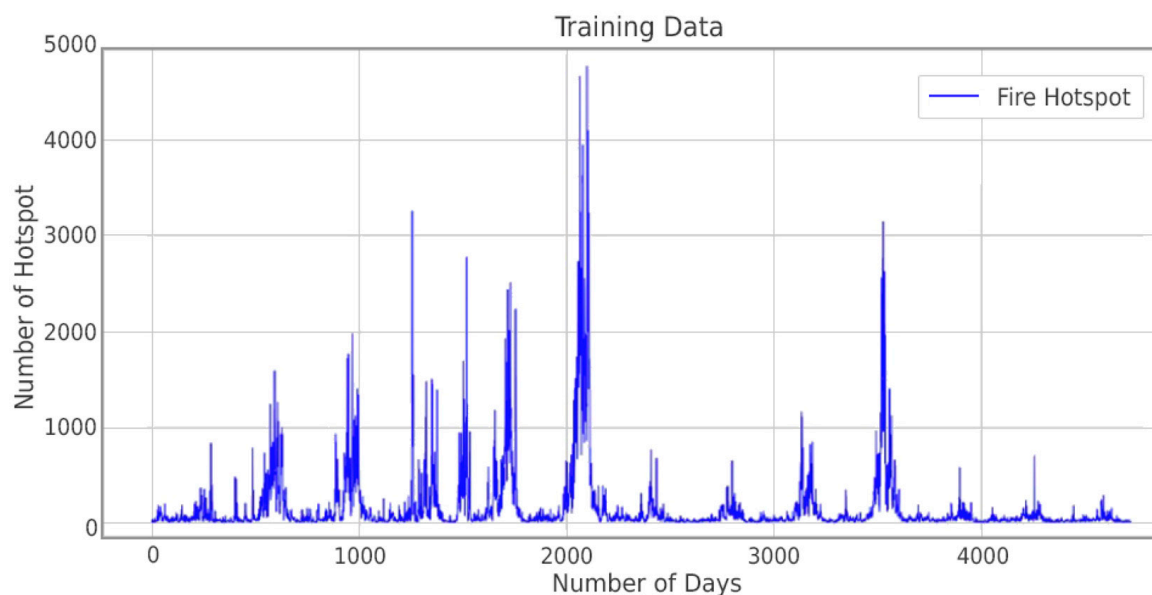**Figure 6.** Mapping of fire hotspots in Indonesia for the year 2022.

Figure 7. Total number of fire hotspots for 2022.

Machine learning is used in this case for data analysis and training, wherein the history of fire hotspot data is divided into (1) testing and (2) training datasets. Normally, the quantity of the training data is larger than the testing data to achieve a high accuracy of the decision. Figure 8 shows a set of training data with a total number of more than 4000 days compared to testing data for forecasting, distributed from the year 2010 to 2022. Referring to data, the year 2019 shows the highest training data, as shown in Figure 9.

Figure 8. Number of training data to number of days.

The results of the LSTM algorithm used for forecasting fire hotspots in Indonesia tested and compared the real available data for the years 2021 and 2022. Early results of forecasting data from 2018 to 2022 were analyzed for forecasting in the years 2021 and 2022. Figure 9 shows the actual fire hotspots data compared to the achieved forecasting for both the years 2021 and 2022. Figure 10a shows the comparison between forecasting and actual data for the year 2021, and Figure 10b shows comparison data for the year 2022. The results show a similar trend and pattern for both results, i.e., actual and forecasting, with an RMSE for the year 2021 of 4.56% and 9.31% for year 2022 and an overall comparison for 2 years with an average is 6.94%. This method was used to check and demonstrate that the proposed algorithm is working fine with a minimum error of less than 10%. Implementation and application of the LSTM algorithm is different from other methods in its ability for deep analysis for the many datasets of fire hotspots, as well as data in time series, as shown previously.

The final results achieved for the forecasted fire hotspot for the coming year 2023 show reliable agreement between forecasting and real data. The process of data forecasting is based on distribution data training more than 4000 datasets like 80% of total data, while the rest is testing data about 20%. The results, as shown in Figure 10, with the real data in the years 2020 to 2022 and then forecasted for the year 2023, obtain a similar pattern and trend as well as the distribution of data and rise at the end of the year. In the year 2022, the trend decreased, and a small number of fire hotspots were detected due to the environmental effect of cool weather and the rainy season, thus reducing the number of fire hotspots. The overall trend and fire event is similar with an increasing number at the end of the year starting in August to September.

The LSTM model in the RNN algorithm has been applied in this analysis, and the results achieve a minimum error in time series of data. The proposed algorithm works fine in predicting and forecasting fire hotspot data for the future, i.e., in 2023. The actual data plotted as a graph in Figure 9 show that the behavior and pattern of data occurrence are similar. With reference to the results, the error bears a low percentage, and the mean proposed algorithm matches to the case in forecasting forest fire hotspots. The result of the forecasting shows a similar pattern to the previous fire hotspots number and trend as well as the monthly distribution. The use of this model yields high accuracy with an error rate of less than 10% compared to other traditional methods. Furthermore, this forest fire analysis and forecasting in the Indonesian region engages only a few researchers and the case categories occupy a specific case, as in tropical regions, as well as the typical land and

contour. Future research will seek to identify and verify the achieved results and compare them with the uncontrolled external factors in this decision. Other parameters that may influence the decision of the system may be considered as well to achieve high accuracy and fast processing of the results.
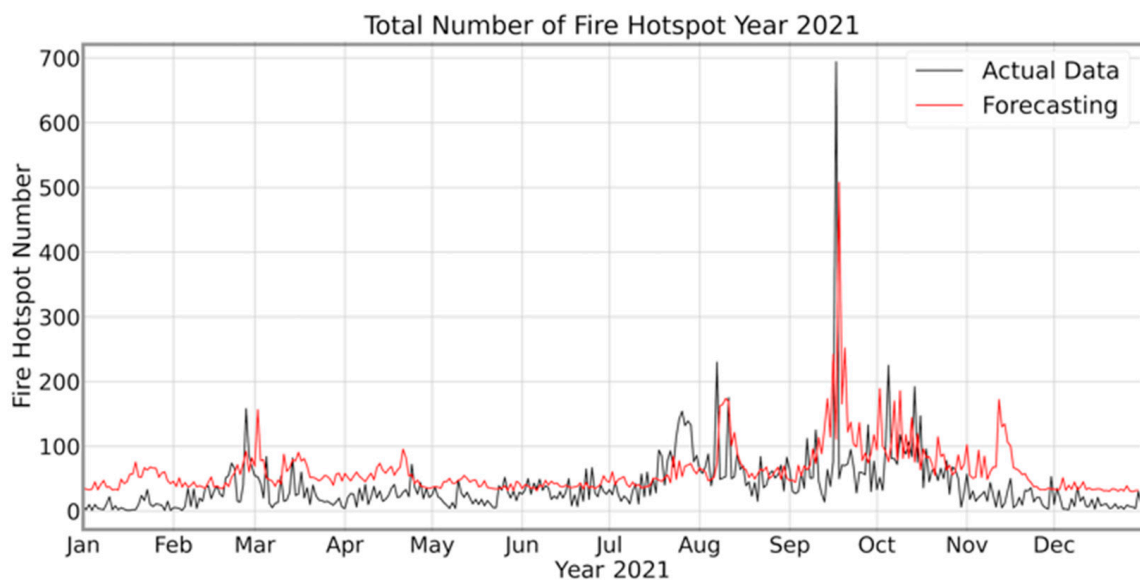
(**a**)
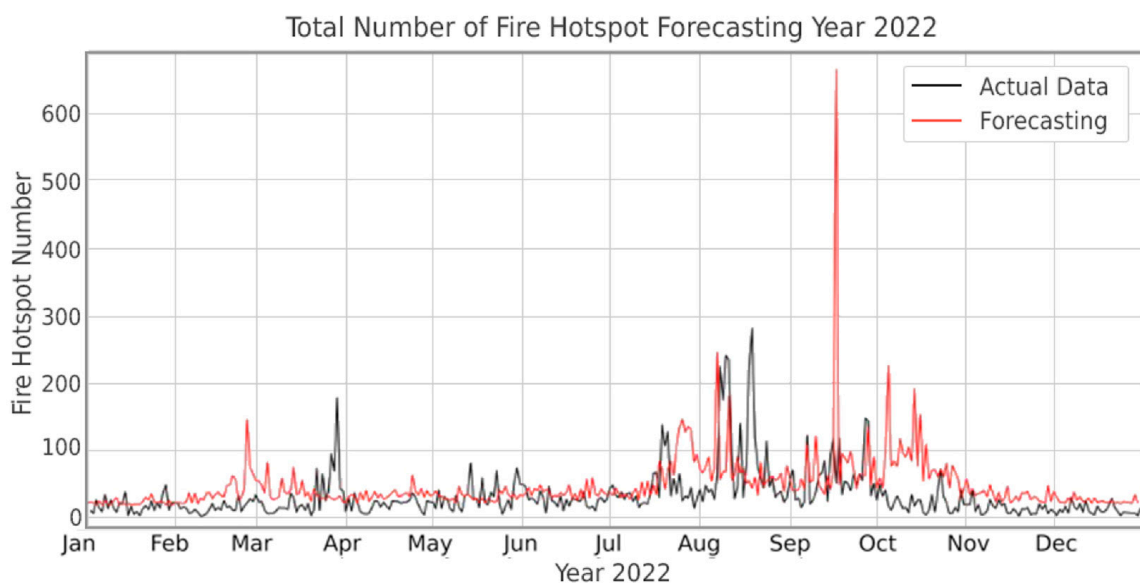
(**b**)

Figure 9. Comparison of actual fire hotspot data and forecasting for the year (**a**) 2021 and (**b**) 2022.

Wildfires pose a significant threat to the ecosystems and communities of Southeast Asia. The region experiences a high occurrence of wild and forest fires, particularly in countries like Indonesia, due to various factors, including climate conditions and natural parameters. Parameters such as climate, wind patterns, and temperature play a crucial role in triggering and exacerbating wildfires in the region. The combination of dry seasons, high temperatures, and strong winds can create favorable conditions for the rapid spread of fires, especially in areas with abundant vegetation and flammable materials. The dry climate contributes to increased vegetation dryness, making it highly susceptible to ignition. Additionally, wind patterns can carry fire embers over vast distances, leading to the rapid

expansion of fire fronts and the ignition of new areas. The interplay of these parameters increases the severity and extent of wildfires, posing significant challenges to firefighting efforts and causing ecological damage, air pollution, and threats to human health. Understanding the impact of these parameters and their interaction is crucial for developing effective strategies for wildfire prevention, early detection, and mitigation in Southeast Asia. The results of this study show the scattering and volume of wildfires in Indonesia, along with forecasting for future years, serving as an alert for the potential occurrence of fires. The work carried out in this study demonstrates an improved understanding of the location and frequency of fire hotspots, particularly in the Indonesia region. The informative graphical representation and future forecasts provided by this study offer valuable insights for addressing wildfire risks.
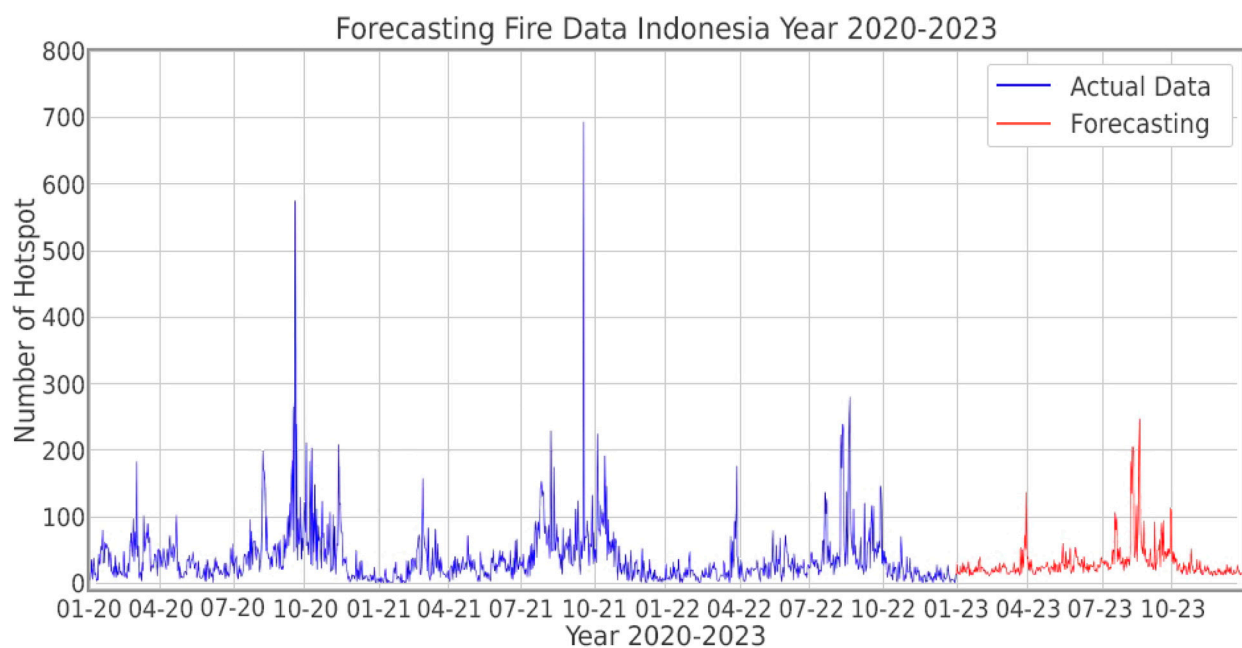
Figure 10. Forecasting of number of fire hotspots for the year 2023.

5. Conclusions

This research discusses the mapping, forecasting, and analysis of fire hotspots in Indonesia. The data have been mapped, revealing that hotspots are concentrated on two major islands, namely Sumatra and Kalimantan. Forecasting has been conducted using NASA's MODIS data, specifically for the Indonesia region, encompassing over 700,000 datasets from 2010 to 2022. The analysis employs the LSTM algorithm, with the data divided into training and testing sets, accounting for 80% and 20%, respectively. The results present a mapping of fire hotspot categories into five levels, differentiating the potential of hotspots to ignite fires. The graph displaying the number of hotspots demonstrates an increasing distribution toward the end of the year, particularly from September to December, coinciding with the dry season or summer. The LSTM algorithm is used for forecasting the year 2023, exhibiting high-performance accuracy with a more than 90% success rate or a mere 6.94% error percentage. In the future, it is crucial to focus on reducing the error percentage by improving the algorithm's training data performance and downsizing the forecasted area to specific states or districts.

Appendix A

Figure A1. Year 2010.

Figure A2. Year 2011.

Figure A3. Year 2012.
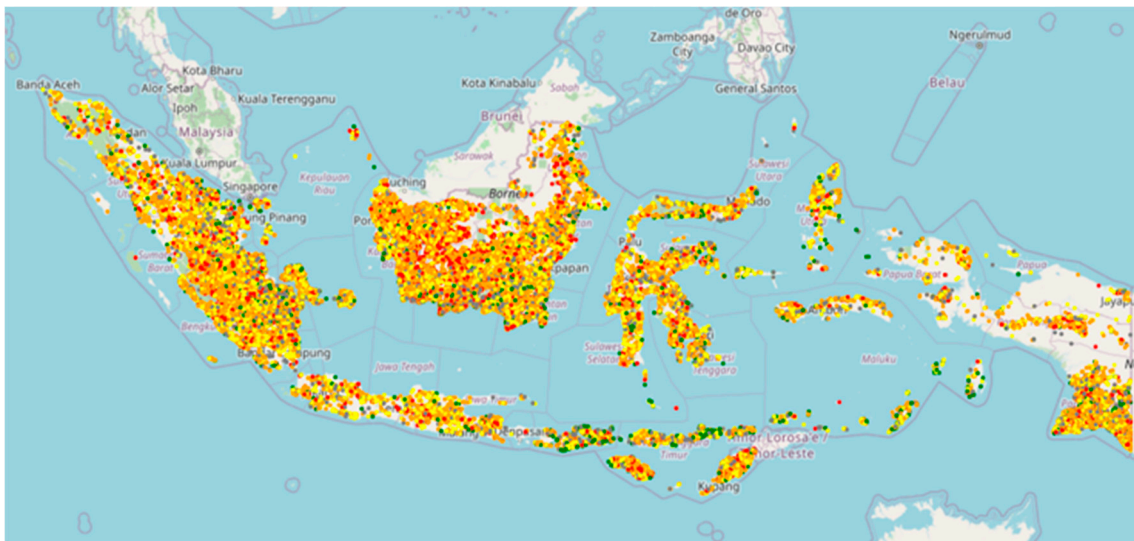

Figure A4. Year 2013.
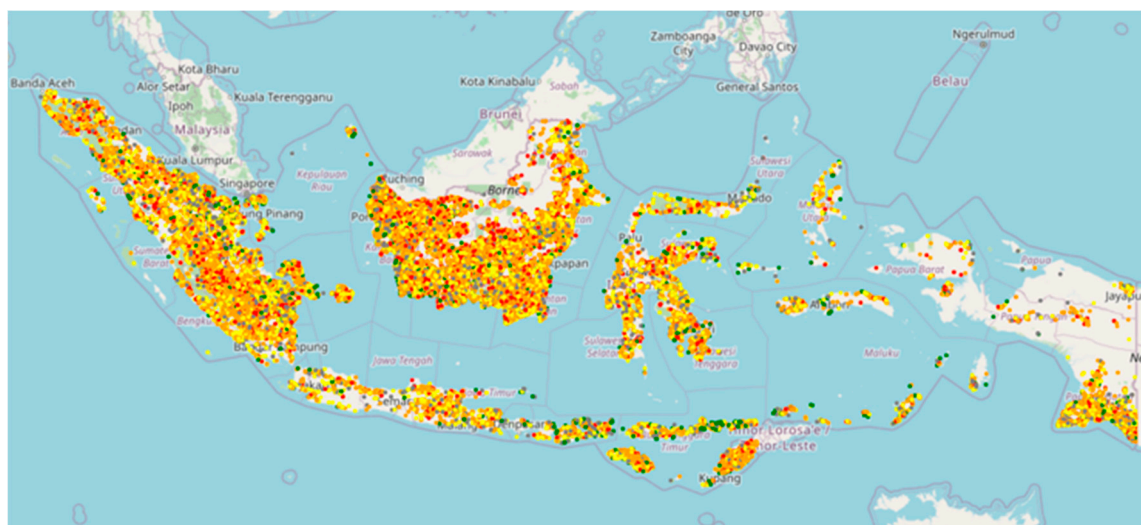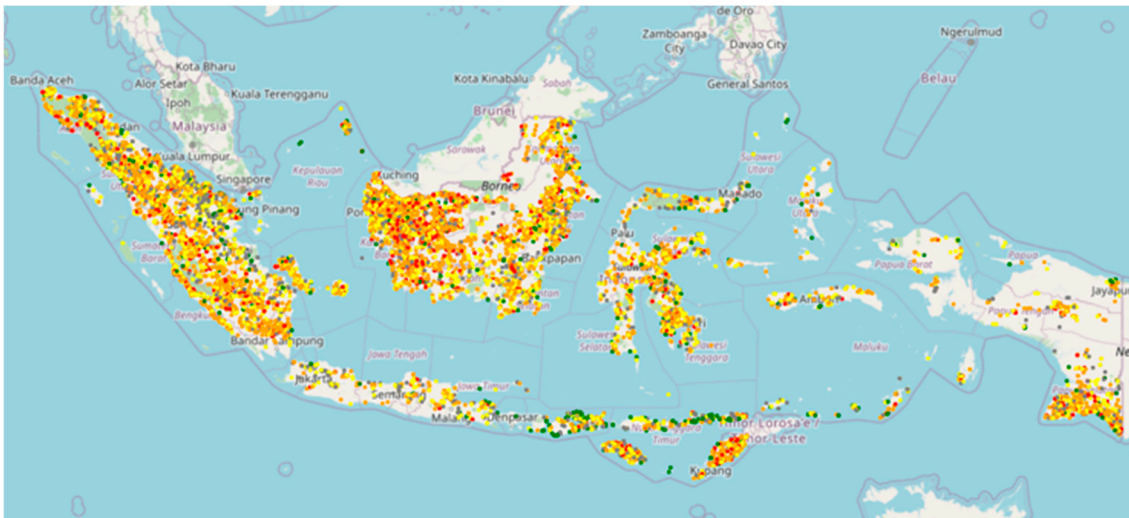

Figure A5. Year 2014.

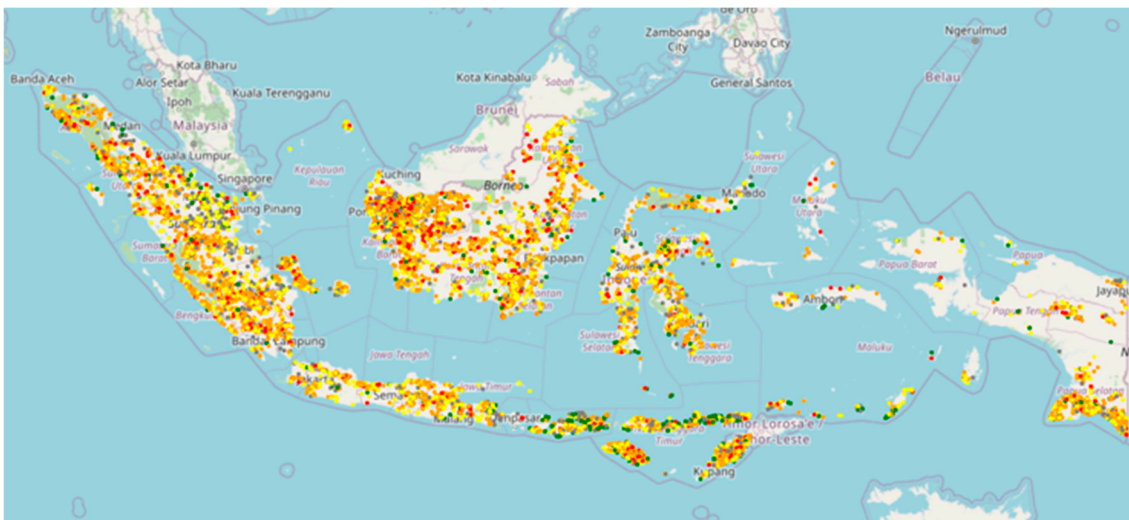**Figure A6.** Year 2015.

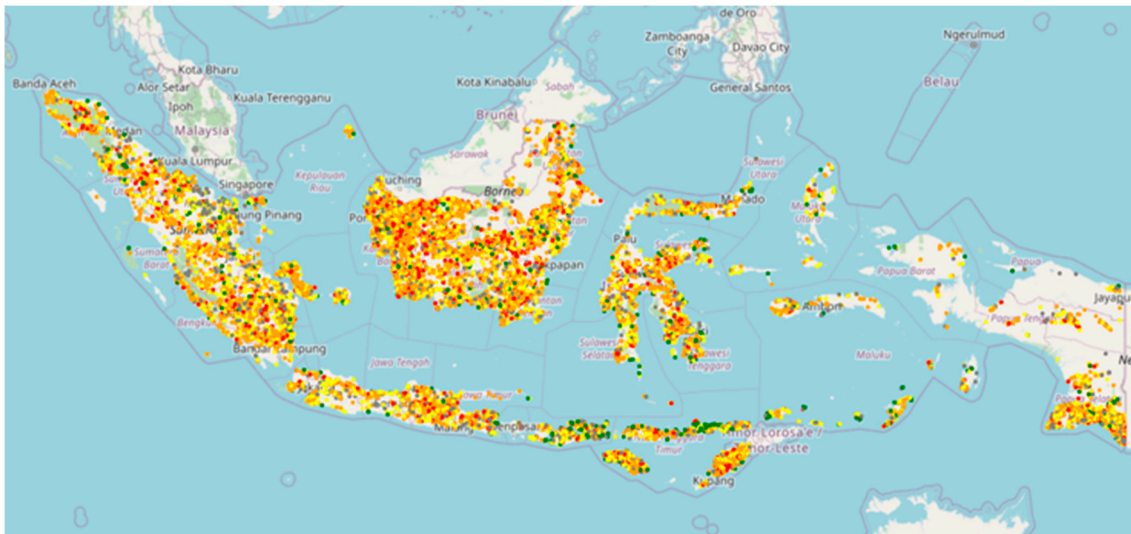**Figure A7.** Year 2016.

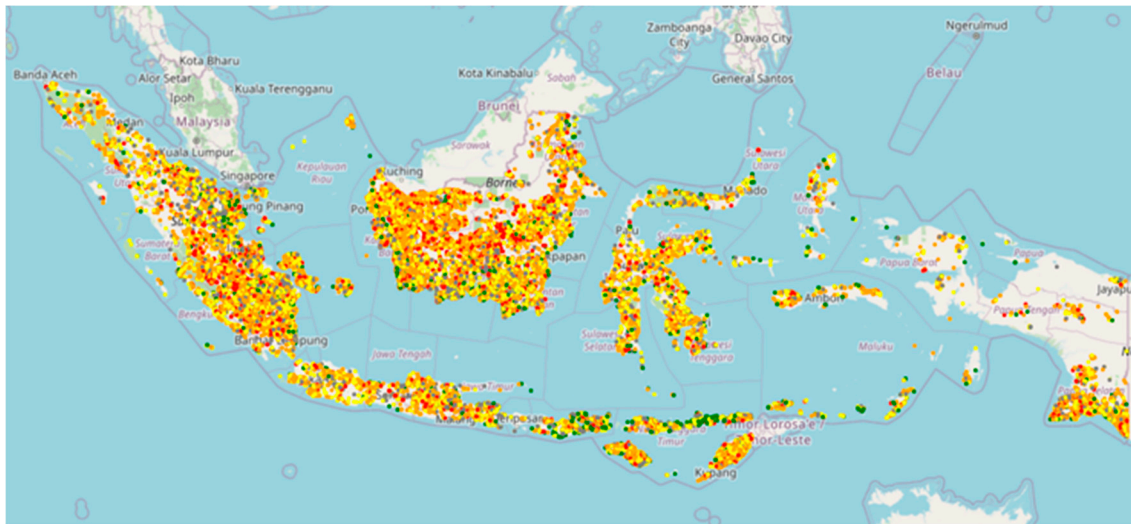**Figure A8.** Year 2017.
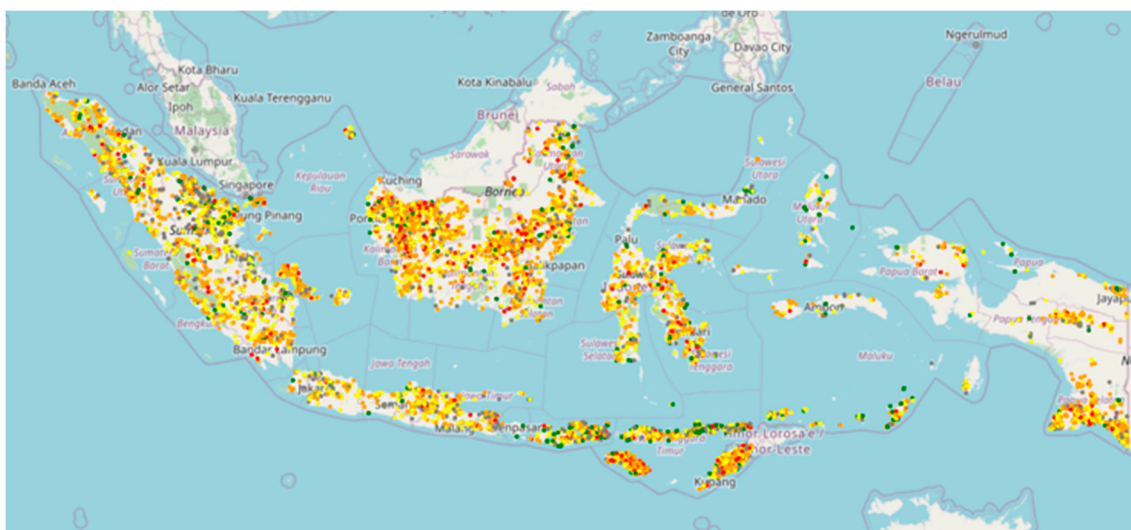
Figure A9. Year 2018.

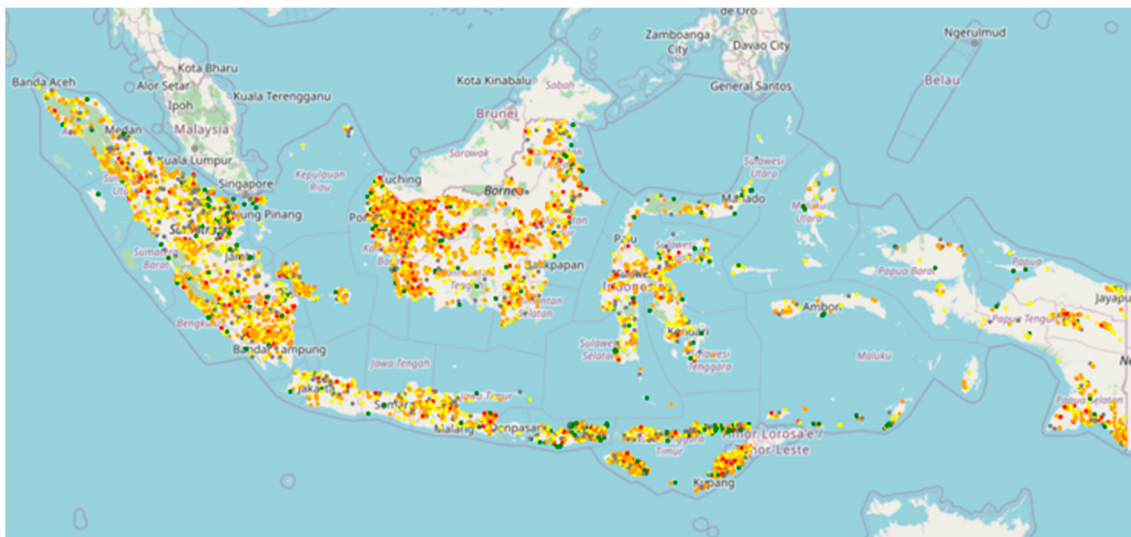
Figure A10. Year 2019.

Figure A11. Year 2020.

Figure A12. Year 2022.

Appendix B
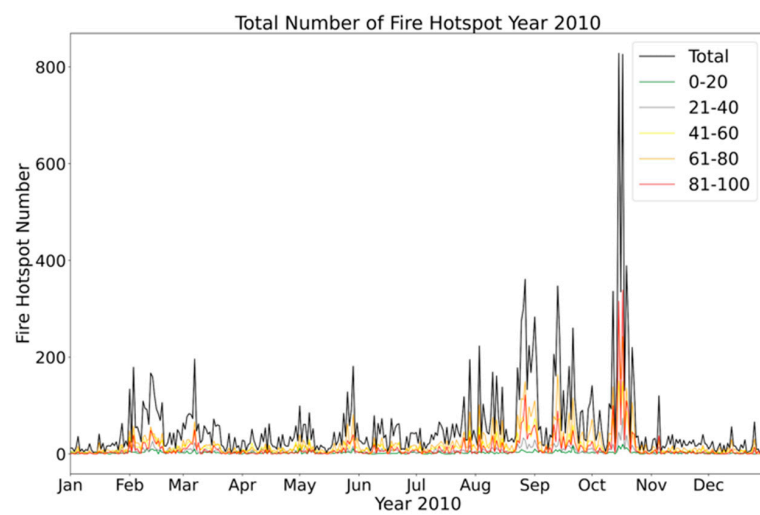
Figure A13. Year 2010.
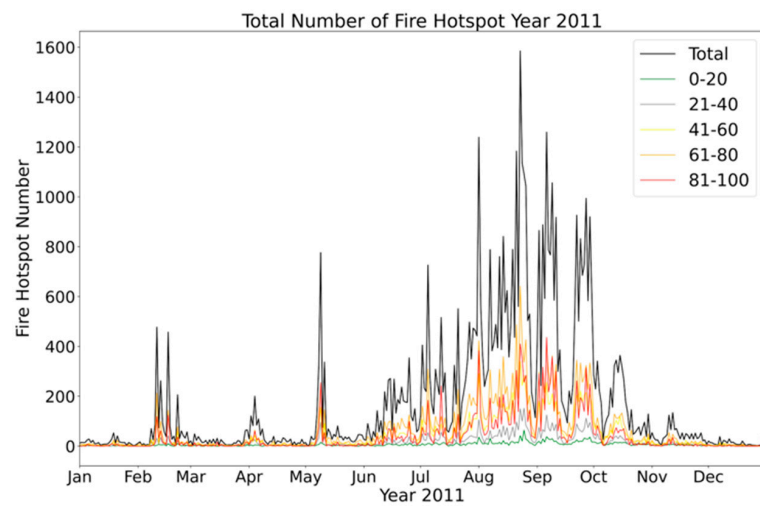
Figure A14. Year 2011.
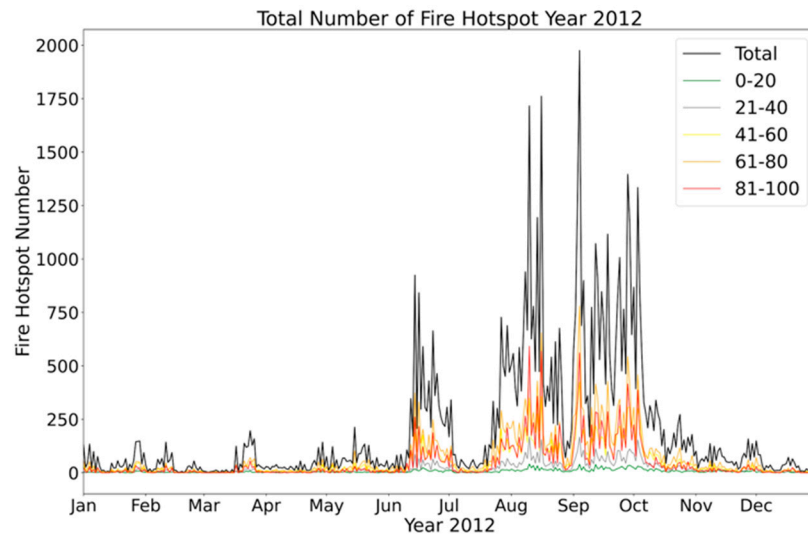
Figure A15. Year 2012.
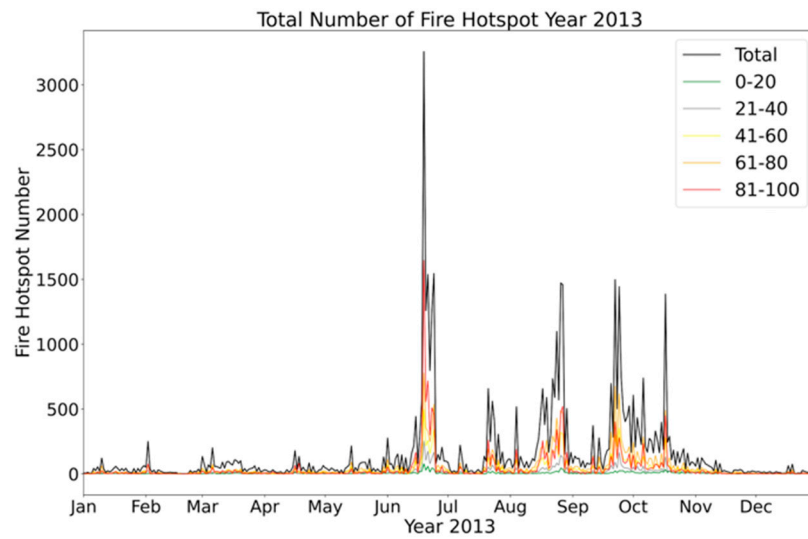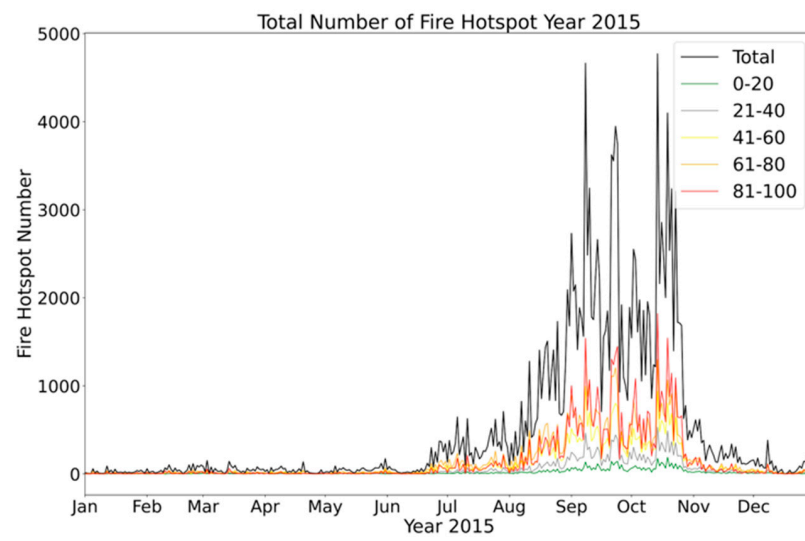

Figure A16. Year 2013.
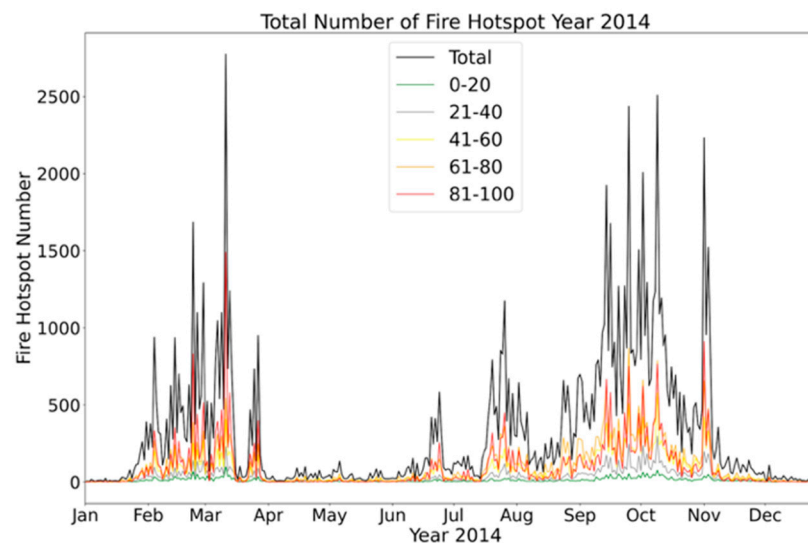
Figure A17. Year 2014.
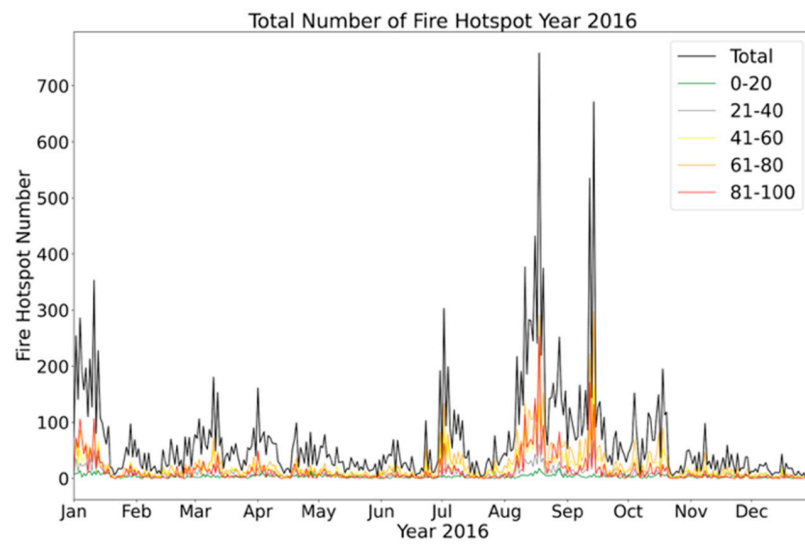
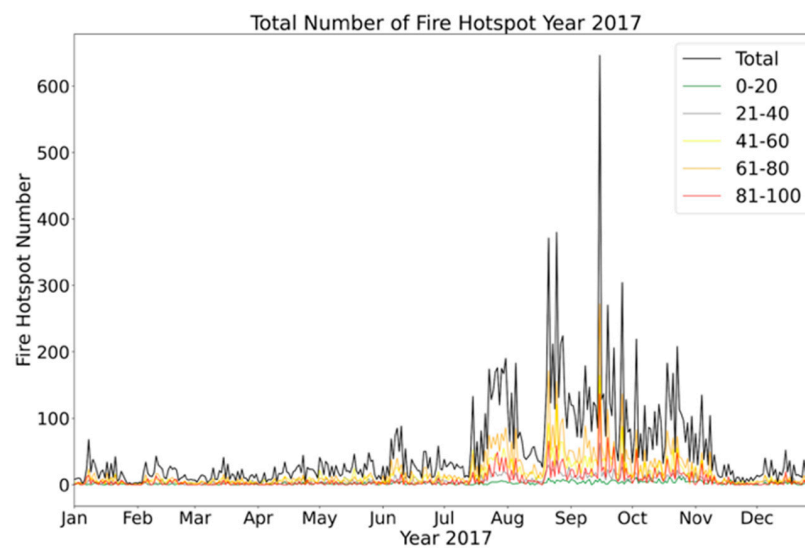**Figure A18.** Year 2015.

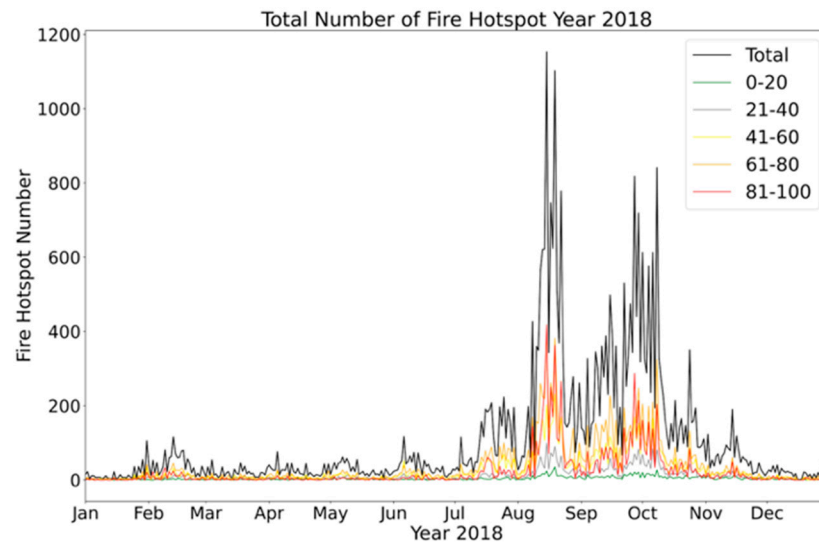**Figure A19.** Year 2016.

Figure A20. Year 2017.
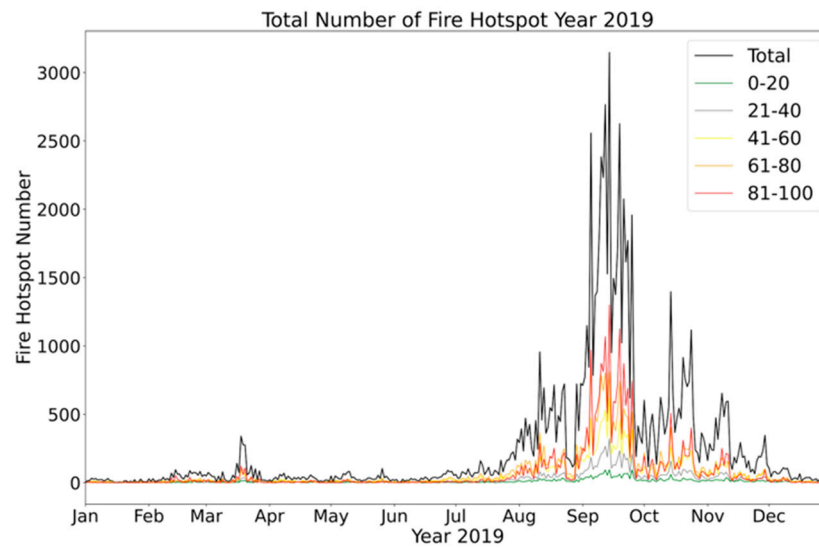
Figure A21. Year 2018.
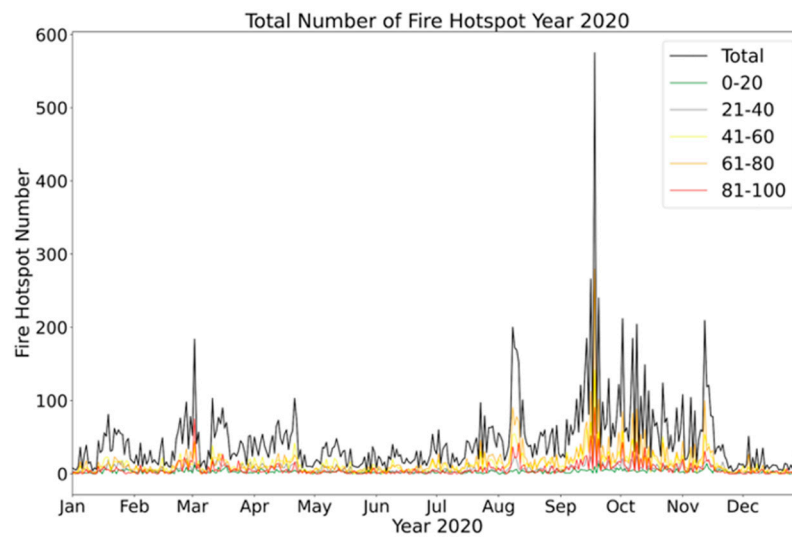

Figure A22. Year 2019.
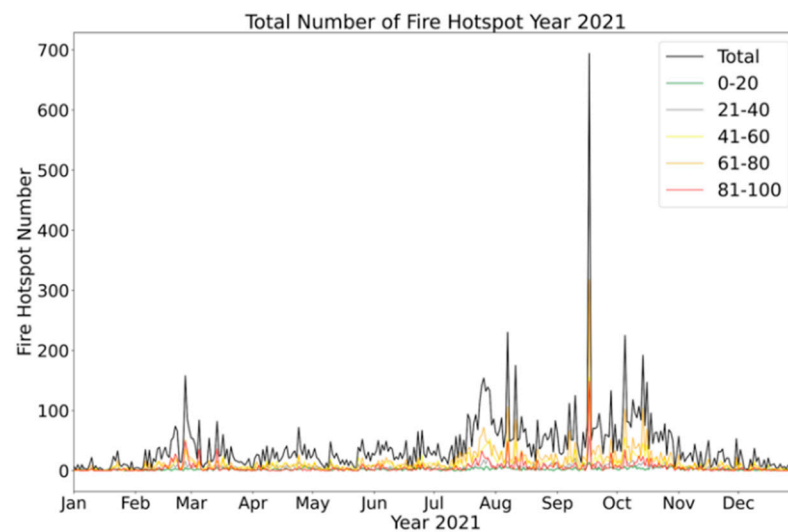

Figure A23. Year 2020.

Figure A24. Year 2021.

References

1. Salavati, G.; Saniei, E.; Ghaderpour, E.; Hassan, Q.K. Wildfire Risk Forecasting Using Weights of Evidence and Statistical Index Models. *Sustainability* **2022**, *14*, 3881. [CrossRef]
2. Dorodnykh, N.; Nikolaychuk, O.; Pestova, J.; Yurin, A. Forest Fire Risk Forecasting with the Aid of Case-Based Reasoning. *Appl. Sci.* **2022**, *12*, 8761. [CrossRef]
3. Lin, X.; Li, Z.; Chen, W.; Sun, X.; Gao, D. Forest Fire Prediction Based on Long- and Short-Term Time-Series Network. *Forests* **2023**, *14*, 778. [CrossRef]
4. Kim, S.-Y.; Muminov, A. Forest Fire Smoke Detection Based on Deep Learning Approaches and Unmanned Aerial Vehicle Images. *Sensors* **2023**, *23*, 5702. [CrossRef] [PubMed]
5. Cheng, S.; Jin, Y.; Harrison, S.P.; Quilodrán-Casas, C.; Prentice, I.C.; Guo, Y.-K.; Arcucci, R. Parameter Flexible Wildfire Prediction Using Machine Learning Techniques: Forward and Inverse Modelling. *Remote Sens.* **2022**, *14*, 3228. [CrossRef]
6. Mohan, K.V.M.; Satish, A.R.; Rao, K.M.; Yarava, R.K.; Babu, G.C. Leveraging Machine Learning to Predict Wild Fires. In Proceedings of the 2021 in 2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 7–9 October 2021; pp. 1393–1400. [CrossRef]
7. Khabarov, N.; Moltchanova, E.; Obersteiner, M. Valuing Weather Observation Systems for Forest Fire Management. *IEEE Syst. J.* **2008**, *2*, 349–357. [CrossRef]
8. Han, J.; Kim, G.; Lee, C.; Han, Y.; Hwang, U.; Kim, S. Predictive Models of Fire via Deep Learning Exploiting Colorific Variation. In Proceedings of the 2019 in International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Okinawa, Japan, 11–13 February 2019; pp. 579–581. [CrossRef]
9. Kadir, E.A.; Irie, H.; Rosa, S.L. Modeling of Wireless Sensor Networks for Detection Land and Forest Fire Hotspot. In Proceedings of the 2019 International Conference on Electronics, Information and Communications (ICEIC), Auckland, New Zealand, 22–25 January 2019; Volume 2019, pp. 1–5. [CrossRef]
10. Nunes, L.J.R.; Raposo, M.A.M.; Meireles, C.I.R.; Pinto Gomes, C.J.; Ribeiro, N.M.C.A. Fire as a Selection Agent for the Dissemination of Invasive Species: Case Study on the Evolution of Forest Coverage. *Environments* **2020**, *7*, 57. [CrossRef]
11. Barmpoutis, P.; Kastridis, A.; Stathaki, T.; Yuan, J.; Shi, M.; Grammalidis, N. Suburban Forest Fire Risk Assessment and Forest Surveillance Using 360-Degree Cameras and a Multiscale Deformable Transformer. *Remote Sens.* **2023**, *15*, 1995. [CrossRef]
12. Kalantar, B.; Ueda, N.; Idrees, M.O.; Janizadeh, S.; Ahmadi, K.; Shabani, F. Forest Fire Susceptibility Prediction Based on Machine Learning Models with Resampling Algorithms on Remote Sensing Data. *Remote Sens.* **2020**, *12*, 3682. [CrossRef]
13. Kukuk, S.B.; Kilimci, Z.H. Comprehensive Analysis of Forest Fire Detection Using Deep Learning Models and Conventional Machine Learning Algorithms. *Int. J. Comput. Exp. Sci. Eng.* **2021**, *7*, 84–94. [CrossRef]
14. Prapas, I.; Kondylatos, S.; Papoutsis, I.; Camps-Valls, G.; Ronco, M.; Fernández-Torres, M.Á.; Guillem, M.P.; Carvalhais, N. Deep Learning Methods for DailyWildfire Danger Forecasting. In Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia, 4 November 2021.
15. Li, X.; Gao, H.; Zhang, M.; Zhang, S.; Gao, Z.; Liu, J.; Sun, S.; Hu, T.; Sun, L. Prediction of Forest Fire Spread Rate Using UAV Images and an LSTM Model Considering the Interaction Between Fire and Wind. *Remote Sens.* **2021**, *13*, 4325. Available online: https://www.mdpi.com/2072-4292/13/21/4325 (accessed on 22 February 2023). [CrossRef]
16. Cao, Y.; Yang, F.; Tang, Q.; Lu, X. An Attention Enhanced Bidirectional LSTM for Early Forest Fire Smoke Recognition. *IEEE Access* **2019**, *7*, 154732–154742. [CrossRef]
17. Natekar, S.; Patil, S.; Nair, A.; Roychowdhury, S. Forest Fire Prediction Using LSTM. In Proceedings of the 2021 in 2nd International Conference for Emerging Technology (INCET), Belagavi, India, 21–23 May 2021; pp. 1–5. [CrossRef]

18. Wang, S.; Hu, Y. A forest fire rescue strategy based on variable extinguishing rate. *Alex. Eng. J.* **2021**, *60*, 1271–1289. [CrossRef]
19. Akther, M.S.; Hassan, Q.K. Remote Sensing-Based Assessment of Fire Danger Conditions Over Boreal Forest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 992–999. [CrossRef]
20. Kadir, E.A.; Othman, M.; Rosa, S.L. Smart Sensor System for Detection and Forecasting Forest Fire Hotspot in Riau Province Indonesia. In Proceedings of the 2021 in International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen, 4–5 July 2021; pp. 1–6. [CrossRef]
21. Zaji, A.H.; Bonakdari, H.; Gharabaghi, B. Remote Sensing Satellite Data Preparation for Simulating and Forecasting River Discharge. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3432–3441. [CrossRef]
22. Costa, R.L.C.; Miranda, E.; Dias, P.; Moreira, J. Experience: Quality Assessment and Improvement on a Forest Fire Dataset. *J. Data Inf. Qual.* **2021**, *13*, 5. [CrossRef]
23. Kadir, E.A.; Rosa, S.L.; Ramadhan, R.A. Detection of Forest Fire Used Multi Sensors System for Peatland Area in Riau Province. In Proceedings of the AIP Conference Proceedings, 2020, Surakarta, Indonesia 6th International Conference on Industrial Mechanical Electrical Chemical Engineering 2020, Surakarta, Indonesia, 17–18 September 2019.
24. Benzekri, W.; El, A.E.; Moussaoui, O.; Berrajaa, M. Early Forest Fire Detection System Using Wireless Sensor Network and Deep Learning. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 496–503. [CrossRef]
25. Varotsos, C.A.; Krapivin, V.F. A new big data approach based on geoecological information-modeling system. *Big Earth Data* **2017**, *1*, 47–63. [CrossRef]
26. Kondratyev, K.Y.; Varotsos, C. Atmospheric greenhouse effect in the context of global climate change. *Nuovo Cim. C* **1995**, *18*, 123–151. [CrossRef]
27. Hartriani, J.; Pusparisa, Y.; Nurhayati, F.; Yudhistira, A.W.; Fitra, S. Forest and land fires over time. *Earth J. Netw.* **2023**, *15*. Available online: https://earthjournalism.net/data-journalism/indonesia-takes-action-but-forest-fires-continue (accessed on 3 March 2023).
28. Kadir, E.A.A.; Rosa, S.L.; Syukur, A.; Othman, M.; Daud, H. Forest Fire Spreading and Carbon Concentration Identification in Tropical Region Indonesia. *Alex. Eng. J.* **2022**, *61*, 1551–1561. [CrossRef]
29. NASA. Fires Hotspot Information. Available online: https://firms.modaps.eosdis.nasa.gov/download/ (accessed on 3 January 2023).
30. Varotsos, C.A.; Cracknell, A.P. Remote Sensing Letters contribution to the success of the Sustainable Development Goals—UN 2030 agenda. *Remote Sens. Lett.* **2020**, *11*, 715–719. [CrossRef]
31. Statista Research Department. Area Burnt by Forest Fires in Indonesia 2015–2021. Available online: https://www.statista.com/statistics/1083530/indonesia-area-burnt-by-forest-fires/ (accessed on 19 January 2023).
32. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
33. Ghosh, R.; Phadikar, S.; Deb, N.; Sinha, N.; Das, P.; Ghaderpour, E. Automatic Eyeblink and Muscular Artifact Detection and Removal from EEG Signals Using k-Nearest Neighbour Classifier and Long Short-Term Memory Networks. *IEEE Sens. J.* **2023**, *23*, 5422–5436. [CrossRef]
34. Le, X.-H.; Ho, H.V.; Lee, G.; Jung, S. Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water* **2019**, *11*, 1387. Available online: https://www.mdpi.com/2073-4441/11/7/1387 (accessed on 8 March 2023). [CrossRef]
35. Chen, H.; Guan, M.; Li, H. Air Quality Prediction Based on Integrated Dual LSTM Model. *IEEE Access* **2021**, *9*, 93285–93297. [CrossRef]