Article

Smoke Detection from Himawari-8 Satellite Data over Kalimantan Island Using Multilayer Perceptrons

Yuhao Mo 1, Xin Yang 1,*, Hong Tang 2 and Zhigang Li 1

1 College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China; 202021490019@mail.bnu.edu.cn (Y.M.); 202131490008@mail.bnu.edu.cn (Z.L.)
2 State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; tanghong@bnu.edu.cn
* Correspondence: yangxin@bnu.edu.cn; Tel.: +86-136-9327-8580

Abstract: The smoke from biomass burning on Kalimantan Island has caused severe environmental problems in Southeast Asia’s primary burning regions and surrounding regions due to the overspread haze. To monitor the biomass burning aerosol plumes on Kalimantan Island, the high-temporal-resolution Himawari-8 satellite data were used in this study. However, studies are limited on smoke detection using satellite remote sensing for Kalimantan Island because of the difficulty caused by frequently occurring clouds and the lack of prior knowledge on applying traditional threshold methods. In this study, we used the multilayer perceptron (MLP) method to identify smoke over Kalimantan Island in August 2015, one of the most severe fire seasons. To prepare sufficient supervision information, a pixel-level labeled dataset was established based on the Himawari-8 data. Based on the labeled dataset, three MLP approaches and two sampling methods were applied to create training samples. A comparison between the detection results for the MLP approaches and classification tree analysis (i.e., CTA) showed that MLP is superior to CTA. The visualization results also showed that the detected smoke areas included those mixed with clouds. Some detected smoke is difficult to identify by the human eye, suggesting that the explanatory dataset built for this study is sufficiently comprehensive. Therefore, the pixel-level labeled dataset and MLP are suitable for regions that are frequently cloud-covered.

Keywords: Himawari-8; smoke detection; multilayer perceptron; pixel classification

1. Introduction

On Kalimantan Island, the peat forests’ frequently occurring biomass burning emits high amounts of aerosol particles into the atmosphere, leading to serious environmental problems [1]. Biomass-burning aerosols also lead to substantial radiative impacts through their interaction with solar radiation [2]. Environmental and radiative effects are not limited to the onset of biomass burning and spread downwind depending on the meteorological conditions. Therefore, detecting the spatial and temporal distribution of biomass-burning aerosols is important for the forecast of air quality and provides insight into the radiative effect of Southeast Asia and its surrounding regions [3].

With the development of remote sensing, smoke detection algorithms have been gradually applied to various sensors. Some researchers have used the images and video sequences from surveillance cameras to identify smoke [4–6]. However, the scope of such identification and monitoring is relatively small and unsuitable for the large areas of smoke emissions from fire. Other researchers have used satellite remote sensing to observe fire hot spots, hence locating nearby smoke plumes. The most widely used optical remote sensing sensors rapidly monitor ground conditions within a large range with low economic cost. These optical remote sensing sensors also contain spectral feature information in a variety of bands. Sun-synchronous orbit satellites observe the same area at the same local time, encompassing the global area. Currently, these types of satellite data have been
widely used in smoke detection [7–9]. However, sun-synchronous satellites only provide observations of transit time. Thus, the variations in smoke in space–time distribution during the remainder of the day can be missed [10]. Therefore, geostationary satellite data have been used to monitor smoke plumes and their diurnal variation in recent years [11–13]. The new-generation geostationary Himawari-8 satellite is equipped with an Advanced Himawari-Imager (AHI) with 16 bands that span from visible to infrared. The AHI provides an image of the Earth’s hemisphere every 10 min with a spatial resolution of 0.5–2 km [14] and facilitates the continuous detection of biomass-burning aerosols.

According to spectral feature analysis [15] and visual analysis [12], optical remote sensing satellite images from the visible, near-infrared and infrared bands contain spectral information from both biomass-burning aerosols and other objects. In a visual analysis of satellite images with visible bands, it is difficult for the human eye to accurately describe the smoke boundary [12]. To fully exploit the geographic and spectral information, diverse methods have been proposed to extract smoke plumes. Previous studies commonly distinguished smoke according to the three RGB (red, green and blue) bands that form a synthesized image [16]. This method achieves an appropriate color display of ground objects for visual interpretation by regulating the band combinations. Although fast and straightforward, this method has low accuracy. Since the band combination and interpretation vary by region, the same RGB band synthesis is not applicable for smoke detection in other regions. To use more spectral information, multi-thresholds concerning the linear relationship among multiple bands classify smoke features [17,18]. For example, Xie et al. [17] proposed a multi-threshold approach to detect smoke in the eastern United States, but the approach could not directly detect the smoke trail. Based on the KMeans algorithm and multiple thresholds, Jing et al. [18] developed a method to distinguish smoke in Heilongjiang (China) that could remove underlaying surface pixels using KMeans to improve the accuracy. Despite balancing computational efficiency and accuracy, these threshold methods require prior knowledge of specific bands and the thresholds among regions for smoke detection. Therefore, these methods are difficult to apply to other satellite datasets. Researchers have begun to explore simplified and adaptable approaches to detecting smoke plumes.

Machine learning algorithms are designed to fit the complicated relationship between input and output and find the optimal function to classify smoke. Current machine learning algorithms, such as Classification Tree Analysis (CTA) [11] and neural network [15,19,20], require many training samples. By contrast, multilayer perceptrons (MLPs) [21] approximate a function from input to output through a neural network. This algorithm type detects smoke pixels through one-dimensional spectral features that consider the spatial features of the smoke. Therefore, MLPs achieve superior results to convolutional neural networks in a relatively small dataset. Some researchers used multi-threshold methods to establish labels for training data through MLPs based on pixel-level classification [15]. Since there is no guarantee of classification accuracy for these label datasets, it is difficult to verify whether the machine learning algorithm can determine the accurate spectral characteristics of smoke. According to the machine learning algorithm, various datasets, including site data [11] and scene classification data [19,20], are used in smoke detection. The site data are relatively more accurate; however, it is often the case that there is an insufficient number of surrounding sites to provide observations as input data. The scene classification method only assigns a category to a scene, leading to uncertainty in the corresponded pixels.

To solve the lack of pixel-level datasets in smoke detection, we established a manually labeled pixel-level dataset. Himawari-8 Level one full-disk images that encompass most of Kalimantan Island on 17, 21, 23, 25 and 30 August 2015 were used in the dataset. The calibrated reflectance and radiance in Himawari-8 Level one are widely used in recent studies [22,23]. The year 2015 was selected because it was the first year in which Himawari-8 provided data, and it was a year marked by high-intensity fire [24,25]. The burning period began in August 2015 during Kalimantan’s rainy season [26]. To verify MLP’s capability in cloudy conditions, we designed three MLP approaches trained by this
dataset. Random sampling and stratified sampling were designed to verify whether the test precision could be improved by changing the ratio between under-sampling non-smoke pixels and oversampling smoke pixels.

The remainder of this paper is structured as follows: Section 2 introduces the study region, dataset and methodology; Section 3 shows the detailed analysis and results; and Section 4 provides further discussion.

2. Study Region, Dataset and Methodology

2.1. Study Area and Dataset

Our study area encompassed most of the Kalimantan Island in Southeast Asia (107.4°E–119.0°E, 5.48°N–4.54°S). The fire hotspots at 05:00 UTC on 30 August 2015 recorded by the Himawari-8 monitor are shown in Figure 1. To improve the contrast, we used a histogram equalization algorithm [27] before visualization. According to Figure 1, the fire hotspots and smoke areas were mainly concentrated in the southeast of Kalimantan Island. The smoke in this area contains many light-absorbing particles and atmospheric plumes with high research value. The peat forest burning event began in August and lasted through to October 2015 [3]. As clouds often cover the study area, we selected images that were representative of the imaging time to avoid overlap with heavy clouds.

![Figure 1. Study area. A histogram equalization of the Himawari-8 AHI image from 30 August 2015 at UTC = 05:00. The wildfire hotspots at UTC = 05:00 on 30 August 2015 are indicated by orange points and the coastline by a red line.](image)

The Himawari-8 satellite is a geostationary meteorological satellite launched by the Japan Meteorological Agency in October 2014 and began service in January 2015, creating images every 10 min. However, there were no observations at 02:40 and 14:40 UTC due to housekeeping. The Level 1 full-disk data contain 16 bands of information including albedo in bands 1–6 and brightness temperature in bands 7–16 [14]. The spatial resolution of bands 1–4 is 0.5–1 km, and that of bands 5–16 is 2 km. We resized bands 1–4 to a 2-kilometer spatial resolution for the convenience of subsequent operations.
To ensure proper image brightness, we studied the sun’s altitude angle when it was most prominent in the daytime. The imaging dates and times of the Himawari-8 Level 1 full-disk data used in this study are shown in Table 1.

**Table 1.** The Himawari-8 Level 1 full-disk data used in the study.

| Imaging Date    | Imaging Time (UTC) | Imaging Time (Local Time) | Number of Images | Pixel Number of Images |
|-----------------|--------------------|--------------------------|------------------|------------------------|
| 17 August 2015  | 01:00–05:00        | 09:00–13:00              | 24               | 6,997,968              |
| 21 August 2015  | 01:00–07:00        | 09:00–15:00              | 36               | 10,496,952             |
| 23 August 2015  | 01:00–04:00        | 09:00–12:00              | 18               | 5,248,476              |
| 25 August 2015  | 01:00–06:00        | 09:00–14:00              | 30               | 8,747,460              |
| 30 August 2015  | 01:00–07:00        | 09:00–15:00              | 36               | 10,496,952             |

2.2. Data Preprocessing

For all the images in Table 1, we manually marked every pixel with a smoke or non-smoke label. This manual interpretation is performed based on the Himawari-8 image and distribution of hotspots, ensuring the consistency of smoke and fire in space and time.

Our marks were the same length and width as the image, dividing every pixel into the following two classes: smoke and non-smoke. Firstly, we created vector data to label every piece of smoke based on every image and transformed the vector data into the raster data whose size is the same as the images. This process increased calculation cost, thus limiting the speed to some extent.

We analyzed the spectral features of the smoke and non-smoke pixels based on marks in all the images in the data set, as shown in Figure 2. After comparing the differences in the mean and variance in the two curves and considering the correlation between long wavebands, we only chose bands 1, 2, 3, 7, 11, 13, 14, 15 and 16 as the input data of our MLP. We used the visible bands (bands 1, 2 and 3) and infrared bands (bands 7, 11, 13, 14, 15 and 16) because most of the objects are distinguished by visible bands and bands 5, 7, 11, 13, 14, 15 and 16 estimate cloud properties [14]. According to Figure 2, band 7 was selected since the difference between smoke pixels and non-smoke pixels was larger than that of band 8.

![Albedo Comparison](image1.png) ![Brightness Temperature Comparison](image2.png)

**Figure 2.** The statistical mean of smoke pixels and non-smoke pixels in shortwave albedo (a) and longwave brightness temperature (b). The X stick represents the bands in AHI. The units of bands 7–16 are K. The orange dots represent smoke pixels and blue dots represent non-smoke pixels. The error bar represents the standard deviation.
2.3. Architecture

A multilayer perceptron is an algorithm based on the perceptron model. It multiplies the nodes of each layer by weight and adds the bias. The weight and bias are determined by the backpropagation loss algorithm, so that the loss of the multilayer perceptron in the sample classification approaches the minimum [21]. After the activation function, it is output to each node of the next layer.

As shown in Figure 3, there were nine nodes in the input layer that corresponded to the value of each band. There were three layers in the hidden layer, with 256 nodes in each layer and two nodes in the output layer. We used the rectified linear unit function (ReLU) in the input layer, the hidden layer and the sigmoid function in the output layer. ReLU is a commonly used nonlinear activation function. It does not change the input’s positive value but renders the input’s negative value to 0. A sigmoid is a classical nonlinear activation function that limits the output of all input values from 0 to 1. ReLU was used in input and hidden layers to prevent over fitting. Sigmoid was used in the output layer so that the output value represented the probability of the pixel for each category.

![Figure 3. Illustration of the multilayer perceptron. The yellow arrow represents nonlinear functions, where the MLP is ReLU(ax+b), the MLP-BN is BN(ReLU(ax+b)) and the MLP-BN-Dropout is Dropout(BN(ReLU(ax+b))), where x is the input of the function, w is the weight and b represents bias.](image-url)

We designed MLP-BN and MLP-BN-Dropout models. The MLP-BN was based on MLP, and the batch normalization layer (BN) [28] was added after each ReLU [29]. The batch normalization layer forcibly stretched the data to the normal distribution with a mean value of 0 and a variance of 1. Then, a linear unit was used to fine tune the distribution of the samples according to the learning situation. The convergence of the network accelerated powerfully following this process. The MLP-BN-Dropout model is based on the MLP-BN layer model, which added a dropout operation after each BN. The dropout layer randomly allowed certain layer nodes to have a weight of 0 in this forward propagation. The weight value in the back propagation was not updated by these nodes. Therefore, this operation can prevent MLP from over fitting.

The formula of batch normalization [28] is as follows:

\[
\mu = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \quad (1)
\]

\[
\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2, \quad \quad (2)
\]
where \( x \) is the input value of batch normalization; \( n \) represents the number of input values; \( \gamma \) and \( \beta \) are the parameters that realize the linear transform; and \( \epsilon \) is \( 1 \times 10^{-7} \), preventing the denominator from becoming 0.

### 2.4. Sampling

We split data imaging from 17, 21, 23 and 25 August 2015 into a training and validation set, leaving data imaging from August 30 for test data. Each image had 291,582 pixels; thus, the total pixel number of 108 images reached 31,490,856. On the one hand, the large number of pixels guaranteed sufficient data for training and testing; on the other hand, they caused computational problems. Directly training all the pixels from each scene image decreased the neural network’s training speed and prevented it from successfully fitting. When compared with the low speed of neural network training, the uneven proportion of smoke pixels and smoke-free pixels was more challenging to solve. The total proportion of smoke pixels was 6‰; therefore, we had to optimize the neural network’s training process by sampling. To ensure the same proportion of pixels from each image in the training dataset, we used the same sampling method for each image. Our experiments suggested that a sampling number near 5000 (such as 4000, 5000 and 6000) pixels [15] in one image had no sufficient representation in our study. Since this study unintentionally provided optimal solutions to balance the quality of the results and computational cost, we obtained a considerable sampling number of 50,000 pixels in each image.

For pixels labeled as non-smoke, we randomly extracted 47,500 pixels without returning the samples. For pixels labeled as smoke, we adopted the same method. However, if the number of smoke pixels in the scene was less than 2500 and the pixels were removed after sampling, we replaced them and performed the sampling method again until the number of accumulated smoke pixels reached 2500. Compared with the direct sampling without replacing the removed smoke pixels, this method ensures that all the smoke pixels are added to the training set when the image’s number of smoke pixels is insufficient at a certain time.

We used the same sampling method as the training set to create the validation set; however, the number of samples was halved. Each image has a sampling of 25,000 pixels and 23,750 smokeless pixels were randomly sampled without being replaced except for the non-smoke pixels sampled for the training set. With the exception of smoke pixels that were sampled for the training set, 1250 smoke pixels were not replaced for sampling. If the number of smoke pixels in the scene was insufficient, all the smoke pixels in the scene image were replaced, and then all the smoke pixels were randomly sampled without being returned until the number of samples reached 1250; this led to the same smoke pixels in the validation and training set, but the small number of the same pixels did not affect the experiment.

The training data set obtained using this sampling method treated the image as the sampling object, artificially increasing the sample proportion of smoke pixels in the training set from 6‰ to 5%. Thus, the neural network learns more about the spectral features of smoke in visible bands. In addition, the sampling proportion and quantity were the same for every image; therefore, the data observed at different times had the same importance in the training sample set. By developing this procedure, we intended to help MLP learn more spectral characteristics of smoke.

To compare with stratified sampling, we designed a random sampling method for this study. We randomly sampled 75,000 pixels in each image, with 50,000 pixels used as samples in the training sets and 25,000 pixels as samples in the validation sets. The training dataset obtained using this random sampling method ensured that the proportion of positive and negative samples was the same as that of four days’ image data. However, the
number of smoke pixels in four days was far less than non-smoke pixels. Random sampling may lead to the overfitting of neural networks and a lack of generalization performance.

2.5. Hyper Parameter

In the training process of the neural network, we used Batch Gradient Descent to organize training data. Each training epoch randomly disturbed the samples in the training set, divided them into the same specifications, placed them into MLP for training in batches and each batch received a loss function. The weight of the neural networks was updated based on the batch of training samples for every backpropagation.

We used cross-entropy loss in our study, and an Adam optimizer, where the learning rate was $3 \times 10^{-4}$, the betas were $[0.9, 0.999]$ and the EPS was $1 \times 10^{-8}$. The parameters described here were the default values in Pytorch. The batch size was 25, the max epoch was 300 and the dropout was 0.5. The process required approximately 10 h to complete all training and the validation process using a V100 graphics card.

3. Results

3.1. Model Evaluation

In the test data set, all the pixels in each image were used to test the generalization performance of MLP. Precision ($P$) and recall ($R$) were used to evaluate the test results of each image. We used macro-precision, macro-recall and macro-F1 to obtain the model performance for all data sets. The calculation formulas are as follows:

$$\text{Macro - precision} = \frac{1}{n} \sum_{i=1}^{n} P_i$$

(5)

$$\text{Macro - recall} = \frac{1}{n} \sum_{i=1}^{n} R_i$$

(6)

$$\text{Macro - F1} = \frac{2 \times \text{Macro - pre} \times \text{Macro - rec}}{\text{Macro - pre} + \text{Macro - rec}}$$

(7)

where $P$ is the precision of each image, $R$ is the recall of each image and $n$ represents the number of images.

Since thin smoke may be accidentally missed when labeling samples manually, there can be many false-positive samples, resulting in a smaller macro-precision. Therefore, we used macro-F1 to find the harmonic average of macro-precision and macro-recall and considered the generalization performance of the model in the test results. The influence of each image on these evaluation indexes was the same.

We compared the MLP based on stratified samples (i.e., MLP-S) and the MLP based on random samples (i.e., MLP-R). In Table 2, the macro-recall of MLP-S is higher than that of MLP-R. Since the stratified samples contained more smoke pixels, MLP-S learned the deep spectral characteristics of smoke more effectively than MLP-R. Therefore, MLP-S recognized more smoke in the test set and increased the number of true positive samples to improve macro-recall. The macro-recall shows that MLP-S could not directly identify the smoke pixels from test sets; therefore, it classified most of the pixels into non-smoke, causing the macro-precision to be high but useless. The low macro-F1 shows that MLP-R could not identify smoke in test sets more accurately. Therefore, in this classification task, stratified samples are superior to random samples.

Verifying the model changes improves generalization ability; therefore, we compared the evaluation indexes of MLP-BN, MLP-BN-Dropout and MLP-S based on the stratified samples. The F1 score in Table 2 shows that MLP-BN-Dropout is the superior of the three model structures, followed by MLP-BN and MLP-S being the worst. These results prove that the BN and dropout layer improved the generalization ability of the model. By comparing the macro-precision and macro-recall of the three models, the BN layer improved the macro-precision from 0.3592 to 0.3967. By contrast, the macro-recall had only a minimal reduction, indicating that the BN layer increased the generalization ability for non-smoke pixels. The MLP-BN-Dropout significantly improved the macro-recall to 0.6884, but that of
MLP-BN only reached 0.5426. The dropout layer significantly improved the generalization ability of the model for smoke pixels because the dropout randomly deleted some nodes in the input and hidden layers. Some spectral features of the samples in the training set could not be learned by the model, preventing the model from overfitting the training set. The dropout alleviated the overfitting and improved the generalization ability of smoke pixels. The BN layer improved the generalization performance in non-smoke pixels.

### Table 2. Classification results of six models on the test data set.

| Model          | Macro-Precision | Macro-Recall | Macro-F1  |
|----------------|-----------------|--------------|-----------|
| MLP-R          | 0.7493          | 0.2197       | 0.3397    |
| MLP-S          | 0.3592          | 0.5568       | 0.4367    |
| MLP-BN         | 0.3967          | 0.5426       | 0.4584    |
| MLP-BN-Dropout | 0.3896          | 0.6884       | 0.4976    |
| CTA-Entropy    | 0.1992          | 0.6040       | 0.2995    |
| CTA-Gini       | 0.2060          | 0.4597       | 0.2845    |

We compared our MLP approaches with the CTA algorithm [11] and used the entropy (CTA-Entropy) and Gini indexes (CTA-Gini) to train in stratified samples. The maximum tree depth was defined as five. The results are shown in the last two lines of Table 2. The macro-precision in these two measures is very low, indicating that smoke-free pixels are rich, and vary with the solar elevation angle. Furthermore, the macro-recall is lower than the MLP-BN-Dropout, showing that the CTA’s performance was worse than the MLP.

Overall, the models based on stratified samples were superior to those based on random samples. Similarly, all the MLP models were superior to the CTA models. The BN layer slightly improved macro-precision, and the dropout layer greatly improved macro-recall. Therefore, MLP-BN-Dropout is superior to the other models.

### 3.2. Visualization Analysis

We visualize the recognition results of three images from the six models in the test set in Figure 4. The three imaging moments of these images correspond to 11:00, 13:00 and 15:00 local time. Clouds increasingly covered the smoke as time passed. As solar radiation weakens and cloud cover increases, the detection of smoke becomes more difficult. The divergence of smoke pixel hues in these three images is a challenge for the model to automatically and accurately identify smoke pixels. We can evaluate the robustness of the models in different situations, including solar radiation and cloud cover.

In the image at 03:00 UTC, there is minimal cloud cover in the southeast of the island; however, the difference between the smoke pixels and clouds is apparent. The MLP-R’s recognition ability in this image is poor. The other three MLP models successfully recognize the outline of smoke in the image. The two CTA models’ recognition results for this image are poor and the pixels near the smoke source are not recognized, on the contrary, many clouds are incorrectly interpreted as smoke.

For the image at 07:00 UTC, the main smoke areas are blocked by clouds, and difficult to identify. The MLP based on the stratified samples still identifies smoke accurately, whereas MLP-R experiences difficulty in identifying smoke pixels. CTA-Entropy overestimates the smoke areas and selects some clouds. CTA-Gini does not recognize the clouds completely and mistakenly classifies them. The results of MLP-BN-Dropout are slightly better than MLP-BN and MLP-S because it can recognize more smoke without mistaking clouds. There is a small smoke area on the sea surface in the northeast region that all the models struggle to identify, because there are few samples at 07:00 UTC, and these methods cannot learn the spectral features of the smoke on the sea surface well. However, the MLP is based on the stratified samples, roughly identifying the smoke area in the northeast of the image at UTC 05:00, whereas the two CTA models cannot recognize smoke well.
Figure 4. The visualization results are based on the histogram equalization images of test data from 30 August 2015. The first line represents the original data image of histogram equalization and the image superimposed with artificial labels (red region). The prediction results of each model are shown from the second line to the last line. The pixels marked in red in the first line represent the original data image of histogram equalization and the image superimposed with artificial labels in the lower six panels are the smoke pixels predicted by the model.

MLP-R recognizes minimal smoke because the basic contour of the smoke is lost. The two CTA models have apparent inaccurate areas and are missing points for images at different times; this shows that the MLP learns more about smoke characteristics from a small number of samples and has a higher generalization ability than the CTA. The results of MLP visualization based on stratified sampling are similar, but the performance is superior to other results. The results of MLP-BN-Dropout are the most accurate and similar to human eye recognition.
Figure 5 shows detailed recognition results of the main smoke areas at 05:00 from Figure 4. Figure 5 shows that MLP-BN-Dropout (b) can distinguish sparse smoke obscured by clouds (as shown in the green box), whereas CTA-Gini (d) mistakenly identifies the thin cloud on the sea as smoke. CTA-Entropy (c) mistakenly identifies the beach and some thin clouds on the coastline as smoke (as shown in the yellow box). As observed in Figure 5, the CTA entropy model is more sensitive to thin smoke, whereas MLP-BN-Dropout tends to identify pixels with high smoke concentration pixels.

**Figure 5.** The detailed visualization results of test data at 05:00 on 30 August 2015, where the red region is the prediction of smoke by models and orange spots are the hot spots.

### 4. Discussion

To solve the problem of missing data with pixel-level biomass-burning aerosol labels, we constructed an artificial label dataset of Kalimantan Island based on Himawari-8 Level one full-disk data. The label dataset contains scenes with a 10-minute resolution, including 01:00 to 07:00 UTC (09:00 to 15:00 local time). We invested our full effort into marking with a higher probability of identifying smoke. It must be ensured that the labels are consistent in context for each pixel. The results of the experiment show that the labeled dataset works effectively in detecting pixels with a biomass-burning aerosol.

We used manually labeled pixel-level datasets to identify smoke, spanning five days of daylight time. The main advantage of our dataset is that it can describe the fire smoke area
of Kalimantan Island in detail and records the short-term change in smoke. Compared with the scene classification label dataset [19], our dataset classifies the smoke more explicitly; thus, models can use the spectral information from each pixel more effectively. Our dataset is more accurate than the threshold segmentation labels [18] and maintains its accuracy under different sun height angles. Compared with the ground site data [11], it has many labels and covers different regions. The labeled dataset based on the geostationary satellite provides near real-time observation of wildfires compared with sun-synchronous satellite data [15,19].

Based on the labeled dataset at the pixel level, we designed an MLP architecture: input layer with nine nodes, three hidden layers with 256 nodes for each layer and an output layer with two nodes. The activation function of the input layer and hidden layers is ReLU, whereas that of the output layer is sigmoid. We created samples with the following two kinds of sampling methods: stratified sampling and random sampling. The trained MLP is based on stratified samples called MLP-S and the MLP trained by random samples called MLP-R. Moreover, we added a BN layer after every ReLU function called MLP-BN and added a Dropout layer after every BN layer called MLP-BN-Dropout. We trained MLP-BN and MLP-BN-Dropout based on stratified samples.

When comparing the MLP-S, MLP-R, MLP-BN and MLP-BN-Dropout, the results show that MLP-S is superior to MLP-R in macro-F1 score and macro-precision. Since we used stratified sampling to improve the proportion of smoke pixels in the whole training set, more spectral features of smoke in visible bands (bands of one, two and three) are learned. The macro-precision of MLP-BN is superior to that of MLP-S because the BN layer causes the features of non-smoke pixels, such as the feature of cloud in band seven, to be easier to determine. The MLP-BN-Dropout shows the highest macro-F1 score because it has the advantage of MLP-BN and prevents the model from overfitting the feature of smoke.

We compared four results from the MLP with two kinds of CTA. The first was based on Gini impurity and another on Shannon entropy [11]. By comparing the prediction results of four MLP methods and two CTA results in Section 3.1, the Macro-F1 score of MLP with a BN layer and a dropout layer is the highest among all the models (0.4976). Compared with the CTA, it has great advantages (CTA-Entropy = 0.2995; CTA-Gini = 0.2845).

The MLP based on stratified samples not only accurately identifies the smoke in the main smoke area but also relatively detects the smoke that is difficult to recognize by the human eye; this suggests that the MLP network is robust in identifying smoke over Kalimantan Island. In the future, we plan on using the high time resolution of the Himawari-8 data and a labeled dataset at the pixel level proposed in this study to monitor smoke in southeastern Asia more efficiently.

Author Contributions: Conceptualization, Y.M. and X.Y.; methodology, Y.M.; software, Y.M.; validation, Y.M. and H.T.; formal analysis, Y.M.; investigation, Y.M.; resources, Y.M.; data curation, Y.M.; writing—original draft preparation, Y.M. and Z.L.; writing—review and editing, Y.M., H.T., X.Y. and Z.L.; visualization, Y.M.; supervision, H.T.; project administration, X.Y.; funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 41971280 and by the Ministry of Science and Technology of China under Grant 2017YFC1501403.

Data Availability Statement: The data are available upon request.

Acknowledgments: The authors would thank for the anonymous reviewers for their helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.
26. Maulud, K.; Fitri, A.; Wan, H.; Wan, S.; Kamarudin, M. A study of spatial and water quality index during dry and rainy seasons at kelantan river basin, peninsular Malaysia. *Arab. J. Geosci.* **2021**, *14*, 85. [CrossRef]

27. Hummel, R. Image enhancement by histogram transformation. *Comput. Graph. Image Process.* **1977**, *6*, 184–195. [CrossRef]

28. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015.

29. Questions About Batch Normalization. Available online: https://forums.fast.ai/t/questions-about-batch-normalization/230 (accessed on 23 November 2016).