

Article

Forecasting Crop Residue Fires in Northeastern China Using Machine Learning

Bing Bai ¹, Hongmei Zhao ^{1,*}, Sumei Zhang ², Xiaolan Li ³, Xuelei Zhang ¹ and Aijun Xiu ¹

¹ Key Laboratory of Wetland Ecology and Environment, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China

² College of Mining Engineering, Taiyuan University of Technology, Taiyuan 030024, China

³ Institute of Atmospheric Environment, China Meteorological Administration, Shenyang 110000, China

* Correspondence: zhaohongmei@iga.ac.cn

Abstract: With repeated changes to local crop residue disposal policies in recent years, the distribution and density of crop residue fire events have been irregular in both space and time. A nonlinear and complex relationship between natural and anthropogenic factors often affects the occurrence of crop residue field fires. To overcome this difficulty, we used the Himawari-8 wildfire data for 2018–2021 to verify the likelihood of crop residue fires against the results of three machine learning methods: logistic regression, backpropagation neural network (BPNN), and decision tree (DT). The results showed the verified accuracies of BPNN and DT methods were 68.59 and 79.59%. Meantime, the sensitivity and specificity of DT performed the best, with the value of area under the curve (AUC) 0.82. Furthermore, among all the influencing factors, open burning prohibition constraints, relative humidity and air pressure showed significant correlations with open burning events. As such, BPNN and DT could accurately forecast the occurrence of agricultural fires. The results presented here may improve the ability to forecast agricultural field fires and provide important advances in understanding fire formation in Northeastern China. They would also provide scientific and technical support for crop fire control and air quality forecasting.



Citation: Bai, B.; Zhao, H.; Zhang, S.; Li, X.; Zhang, X.; Xiu, A. Forecasting Crop Residue Fires in Northeastern China Using Machine Learning. *Atmosphere* **2022**, *13*, 1616. <https://doi.org/10.3390/atmos13101616>

Academic Editors: Miae Kim and Jan Cermak

Received: 26 August 2022

Accepted: 30 September 2022

Published: 3 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: biomass burning; fire forecasting; logistic regression (LR); backpropagation neural network (BPNN); decision tree (DT); artificial intelligence; agricultural fire; policy

1. Introduction

The combustion of crop residues emits carbon dioxide, carbon monoxide, nitric oxide, nitrous oxide, and aerosol particles. During the combustion process, large volumes of gases and particulate matter (PM) may be released in a short period of time, which affect regional haze and cause environmental pollution. In particular, the inhalation of PM causes cardiovascular and respiratory diseases in humans, and has additional adverse effects on human health, atmospheric visibility, and the Earth's radiation balance [1]. Therefore, the burning of agricultural residues is a significant issue in China and is the subject of much active research.

In 2010, around 20.8% of straw fields in China experienced open burning [2], with these events being a major contribution to the total annual amount of biomass that was burned that year. Crop residue open burning can be significantly affected by human activity and weather conditions, making the likelihood of combustion highly uncertain and difficult to forecast. At this time, agricultural residue field burning events are primarily monitored via satellite-based remote sensing data or field observation; however, both of these methods have a lag time [3], which limits their ability to prevent and control crop residue open burning events.

Most previous forecasting studies of biomass burning events, whether in China or elsewhere, have mainly considered fire occurrence, susceptibility, and fire risk in forested regions [4–7]. Forecasting crop residue burning in open regions remains limited because of

the unpredictability and non-linear relationships that exist between natural and anthropogenic factors. It is thus important to accurately quantify how natural factors, human activity, and fire relate to each other [8]. Machine learning (ML) techniques can significantly improve our understanding of the likelihood of fire events, as the fundamental goal of ML is to extract useful knowledge, information, or correlations between datasets in order to use the discovered patterns to predict future events of interest [9,10]. Common ML techniques include logistic regression (LR), artificial neural network (ANN), decision tree (DT) analysis, support-vector machines and deep learning. Of these, LR is powerful for predictability studies, as it may readily handle data with both normal and linear relationships. Further, ANN techniques have many forms, including backpropagation neural network (BPNN), Hopfield neural network, and bidirectional associative memory neural network types. The BPNN type is the most common of these as it considers errors for individual data points [11]. Finally, DT is a non-parametric technique that can automatically select the most important variables within a set; thus, when compared with LR, DT techniques are more flexible at assessing complicated interactions among different variables. While deep learning is also highly capable of learning, its complexity also leads to a dramatic increase in the time and sophistication of the algorithms, requiring higher programming skills and better hardware support [9].

There are some studies based on neural networks to forecast fires. Most of the previous studies focused on forest fires. Up to now, little attention has been paid on cropland fire; even cropland fire is an important type of biomass burning in some Asian regions. Different from forest fires, cropland fires are influenced not only by natural factors, but also by human interference. Therefore, the forecasting of cropland fire would be more complicated. Di Giuseppe [12] predicted straw burning in 2015 in Indonesia and Alaska by adjusting the relationship between the systematic function of forest fire risk climate index and the CAMS model, the human factors and weather conditions that have a large impact on straw open burning were not reflected during the study, and there was much space for improvement. Based on the advantages of ML, ML has been applied to the study of forest wildfire science and management since the 1990s [10]. In China, Feng [13] utilized neural networks to verify the number of crop residue fire points in Southern China during 2013–2015, correlation coefficients calculated with observational data were in the range 0.6–0.8. Our team previously used BPNN to forecast farmland fire points over Songnen Plain in Northeastern China with a forecasting accuracy of 69.7% [14], although the two studies did not consider the effects of anthropogenic factors. The management and control policies of crop residue open burning have changed many times in recent years, particularly in Northeastern China. In 2018, the Straw Open Burning Prohibition Plan was enacted and subsequently enforced across Jilin Province by the Ecology and Environment Department of Jilin Province. This legislation clearly specifies restrictions on straw combustion, including demarcation of areas where open burning straw is prohibited—referred to as straw open burning prohibition areas, and those where combustion is restricted—referred to as prescribed open burning areas (OPEA) for straw [15]. Implementation of these policies represents a major challenge for forecasting crop residue fire events. Previously, we considered anthropogenic factors and applied the BPNN method to verify crop residue fire events across Northeastern China in 2020, which produced a verification accuracy of 60% [16], although there remains much opportunity to improve this prediction technique.

In this study, we considered natural and anthropogenic factors in applying three discrete ML techniques (LR, BPNN, and DT) to verify the probability of crop residue open burning occurrences and their spatial distribution in Northeastern China from 2021, as constrained by satellite-based remote sensing data of fire events. We proposed that if the final verification accuracy from our previous study (60%) increased by at least 10%, to 70%, then this new model would be acceptable. Thus, we investigated which of the three above-mentioned ML methods was most suitable to verify crop residue fire events in Northeastern China when considering natural and anthropogenic factors. The innovation of this study is to integrate the complicated anthropogenic and conventional natural factors

to explore the mechanism of cropland fire occurrence. The results presented here may provide important advances in understanding fire formation in Northeastern China, enhance the capacity for forecasting crop residue field burning events, and the evaluation of fire occurrence in other regions at this scale. At the same time, appropriate preventive measures can be provided to local government regulators and other decision makers in controlling open agriculture burning, especially by designing agricultural waste management strategies that are appropriate for different situations in different regions. Finally, the prediction results of fire emissions can be applied to scale-model transformed air quality models, providing improved data for these atmospheric models and improving their predictive capabilities and early warning systems.

2. Research area and Methodology

2.1. Spatiotemporal Characteristics of the Research Area

Northeastern China ($38^{\circ}42'–53^{\circ}35' N$ and $115^{\circ}32'–135^{\circ}09' E$) includes the Heilongjiang, Jilin, Liaoning, and eastern Inner Mongolia Provinces (Figure 1). This region covers $1,240,000 \text{ km}^2$, has an annual average temperature range of $-1.3^{\circ}C$ to $6.6^{\circ}C$, and an annual total precipitation range of $500–1000 \text{ mm}$ [17]. Northeastern China contains temperate and warm temperate zones that are characterized by a continental monsoonal climate, high temperatures, abundant precipitation, and sunshine during the summer. Soil in fields in this region is mainly Phaeozem and Chernozem, making northeastern China one of the most fertile regions in the world [18,19], with agricultural land covering approximately $40,000 \text{ km}^2$. This advantageous geographical environment and the favorable natural resources promote vegetation growth; thus, many crops are planted each year and the area represents an important region for grain production in China. Northeastern China is also suitable for single-crop farming between May and September each year, with major crops in the region including rice, corn, soybeans, and wheat [19]. In 2015, the crop residue yield from the Northeastern China was 16 billion tons yearly, which accounts for one-fifth of the total crop residue production throughout China; however, the overall utilization ratio in the region is 80%, placing it slightly below the national average utilization level of 83% [20]. Open burning of crop residue is widespread in Northeastern China. Crop open burning events in the Liaoning, Jilin, and Heilongjiang Provinces represented 57.37% of fires across all of China between 2014 and 2018 [21]. In response, Jilin Province prohibited open burning of straw in specific areas from 2018, which included those surrounding cities, highways, railways, national roads, provincial roads, and airports. As a result, the spatial and temporal distribution of agricultural residue open burning is constantly changing.

The main annual growth cycle of crops in Northeastern China involves sowing in spring and harvesting in autumn. Farmers often openly burn surplus crop residue in the field in which they grew to ensure unimpeded sowing in the following spring. According to changes in the daily distribution of field fire events in Northeastern China, most major burning periods occur between 21 March and 15 April of each year. We selected buffers of five days before and after this range to account for the occurrence of anomalous events, such that the periods chosen for study were from 16 March to 20 April each year between 2018 and 2021.

The Himawari-8 satellite data used in this study recorded 31,944 fire points representing open burning of crop residues in the periods of 16 March to 20 April each year between 2018 and 2021. In general, bagging, boosting and fixed data size on the data prevent an overfitting learning phenomenon occurring. In this study, equivalent data with no fires were randomly selected to be added to the data for fire points, which produced an original data set [8,22], as shown in Table 1.

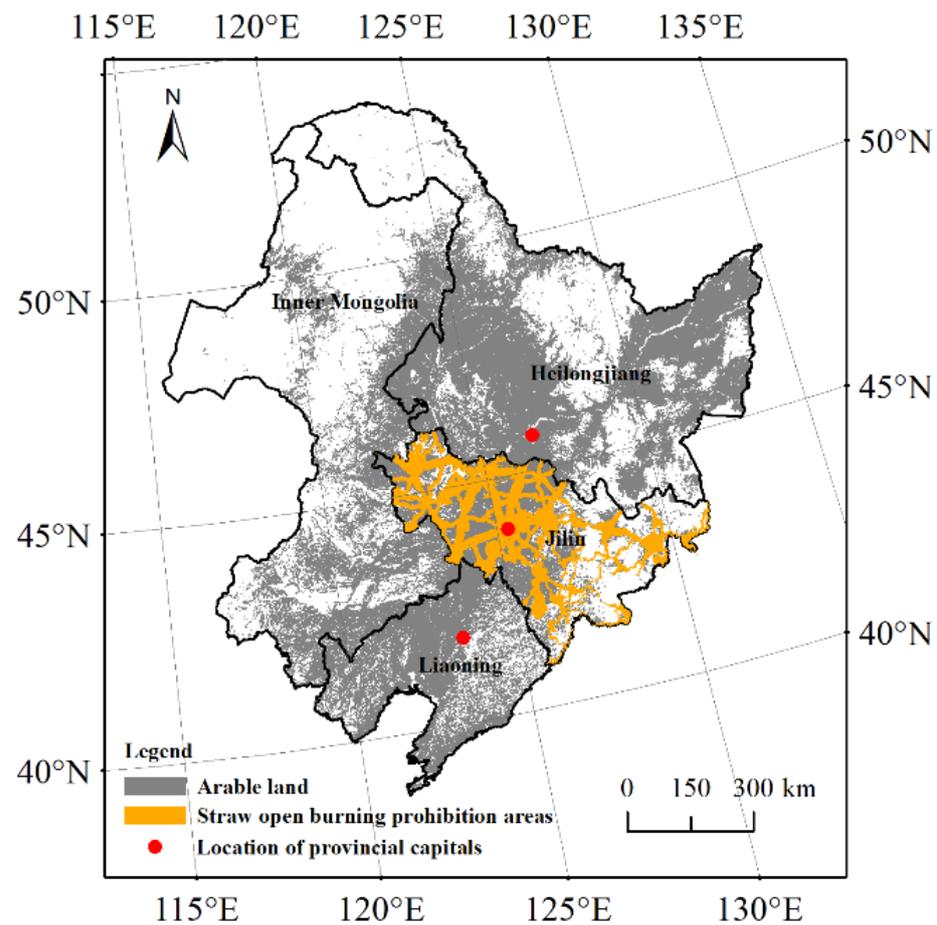


Figure 1. Northeastern location map of China discussed in this study.

Table 1. Detailed information about the input data used in this study.

Training Time	Training Samples	Verifying Time	Verifying Samples	Consideration Variables	Input Variables
16 March to 20 April, 2018–2020	14,924	16 March to 20 April 2021	48,964	Meteorological elements (5), soil moisture content (2), open burning ban policy	WIN, TEM, PRS, PRE, PHU, SOIL, D2-D1, OPEA

WIN—daily maximum wind speed; TEM—mean daily air temperature; PRS—daily mean air pressure; PRE—cumulative 24 h precipitation; PHU—mean daily relative humidity; SOIL—soil moisture content on the day; D2-D1—change in soil moisture content over a period of 24 h; and OPEA—open burning prohibition areas.

2.2. Study Data

2.2.1. Crop Residue Open Burning Events Data

Data about daily fire events’ locations were downloaded from the DWLFBet of Global Daily Fire Location Product [23] during March and April for 2018, 2019, 2020 and 2021. These data were collected by the JAXA Himawari-8 monitor satellite, which has a spatial resolution of 2 km. Many satellites monitor fire point data across China; however, the sensor within the Himawari-8/9 Advanced Himawari Imager employs a full-disk scanning observation method, which has 16 observation bands and performs faster revisit times (around 10 min for a full disk). These properties facilitate the identification and follow-up of rapidly changing weather conditions, and the deduction of quantitative products [24]. The downloaded data included the date, longitude, latitude, and radiative power of fire events. The data were supplied in CSV files, and daily field fire points in Northeastern China were extracted by ArcGIS.

2.2.2. Surface Meteorological Data

Surface meteorological data for the day from March to April each year during 2018–2021 were obtained from the China Meteorological Data Network [25]. Atmospheric precipitation will certainly directly affect the open burning of cropland fires, the relative humidity and wind speed in the air may also determine whether farmers burn crop residue. Furthermore, those parameter changes in relative humidity will affect temperature and pressure. Therefore, five meteorological variables of 839 meteorology stations were selected: daily average air temperature (TEM), daily average wind speed (WIN), daily relative humidity (PHU), 24 h cumulative rainfall (PRE), and daily average air pressure (PRS). Data files were provided in CSV form and were presented in text format. We used ArcGIS to extract information about meteorological stations in Northeastern China. In order to facilitate compatibility of our results with other national and international air quality modellings, surface meteorological data were obtained with a spatial resolution of $3 \text{ km} \times 3 \text{ km}$ using an ordinary kriging method.

2.2.3. Soil Moisture Content Data

At higher soil moisture levels, crop residue open burning events are less likely to occur as wet farmland is unable to conduct farmland activity, therefore, we selected soil moisture content as an influencing factor affecting agricultural fire burning. Soil moisture content data during 2018–2021 were received from the soil moisture project of the European Space Agency's (ESA) Climate Change Initiative (CCI). The ESA CCI SM v06.1 [26] is the most recent product and includes three surface soil moisture datasets: an active product, a passive product, and a combined (active + passive) product. Among these, the active product is most appropriate for densely vegetated areas, while the passive product can better detect fire events in sparsely vegetated regions [27]. Consequently, we used the combined product so that all areas could be analyzed. The global-scale spatial resolution of the combined product is 0.25° . The information was provided in NetCDF-4 file, and data had units of $\text{m}^3 \text{ m}^{-3}$, which represent an absolute value of unit volume. After performing interpolation using an ordinary kriging method and extraction by ArcGIS, we obtained a soil moisture data set for Northeastern China with a spatial resolution of $3 \text{ km} \times 3 \text{ km}$.

2.2.4. Anthropogenic Management and Control Policy

The designation of a straw open burning prohibition area directly affects whether burning will happen in that area; divisional information regarding areas in Jilin Province where burning is prohibited was received from the People's Government of Jilin Province [28]. The data documents were available as vector, and prohibition messages were calculated every 3 km using the 'convert to points' feature of ArcGIS. This spatial information was then imposed onto a map of farmland in Northeastern China. In this way, each grid included the necessary information to know whether open crop residue field burning was permitted. The number of fire points per forbidden area for straw field burning was fixed as 0.

2.3. Methodology

2.3.1. Data Preprocessing

Farmland in Northeastern China was divided into a grid with a spatial resolution of $3 \text{ km} \times 3 \text{ km}$, which produced 78,694 areas. Each area contained information about its natural and anthropogenic characteristics, alongside whether crop residues were burnt in the field. This allowed the following data to be used as inputs to the ML process: five meteorological variables (TEM, WIN, PHU, PRS, and PRE), soil moisture content on the day (SOIL), the change in soil moisture content over a period of 24 h (D2-D1), anthropogenic management and control policy (OPEA), and the occurrence of fire. Of these parameters, the occurrence of fire was taken as the output target. The dependent variable in this case was the field fire, which is defined as true (1) if at least one fire occurred and false (0) if there is no fire in the 24 h interval. Detailed information about the input data used in this

study is shown in Table 1. Three different ML analysis methods were applied by using the SPSS Modeler 14.1 Software: LR, BPNN, and DT.

Considering the accuracy and efficiency of ML training, we normalized the input data as follows [6,8]:

$$x = (x_i - x_{min}) / (x_{max} - x_{min}) \quad (1)$$

where x_i represents a raw data, x_{min} shows the minimum value of raw data, x_{max} means the maximum value of raw data, and x is the input value for normalization.

2.3.2. Logistic Regression (LR)

An LR model is often used to describe the relationship between binary variables (i.e., dependent variables have values of either 0 or 1) and one or more independent variables. In 1993, an LR model was first used to identify areas in San Bernardino National Forest, California, with a high likelihood of fire based on environmental, human, and spatial factors [29]. Subsequently, LR models began to be widely used for research in various fields that required forecasting. In recent decades, many studies around the world have used LR technology to forecast the probability of forest fires [30–32]. The LR method is expressed as:

$$\text{logit}(P_i) = \ln(P_i/1 - P_i) = \alpha_0 + \alpha_1x_1 + \alpha_2x_2 + \dots + \alpha_nx_n \quad (2)$$

in which P is the occurrence probability of burning, with n is for the number of covariates, $(\alpha_1, \alpha_2, \dots, \alpha_n)$ are the coefficients of every variable, that (x_1, x_2, \dots, x_n) are factors that affect whether a fire occurs. We chose the binomial and backward method to build the model, using a maximum of 20 iterations.

2.3.3. Backpropagation Neural Network (BPNN)

A BPNN model is composed of a 3-layer structure, which are the input, hidden and output layers [13,33]. A total of eight types of environmental and anthropogenic information (including WIN, TEM, PRS, PRE, PHU, SOIL, D2-D1 and OPEA) were selected as the input layer data. To define a hidden layer, it is most important to determine the number of layers in the BPNN, and then define two essential parameters: the number of neurons and iterations. These parameters must be determined experimentally to optimize the performance of BPNN modeling, such as minimizing the computational cost and improving the level of test performance [4,13,33]. Through repeated trials, this study setup the hidden layer as a monolayer and simultaneously 20 neurons with 200 iterations. The output layer indicated whether a fire occurred, with results being either 0 (no) or 1 (yes).

2.3.4. Decision Tree (DT)

The DT is a graphic method of intuitively using probability analysis with a tree structure. Each classification tree is generated by bootstrapping samples and uses two-thirds of the samples for training and another one-third to validate. Each partition of the tree is identified via the use of a stochastic subset of forecasters at nodes. The end outcome takes the mean value of all the numbers [34,35]. Because the DT method can measure each variable's importance, the outcomes can be compared to other regression techniques [36]. In this study, the C5.0 algorithm was chosen to build a decision tree model, and we applied 10 boosting tests and 10-fold cross validations to the data. Meanwhile, global pruning was used to reduce errors in the modeling process; here, the pruning purity was 75 and there were at least two records in each branch.

2.3.5. Model Evaluation

This study considered four different scenarios to evaluate the verification accuracy of each model: the verification and the observation both show combustion (TP); the verification and the observation both show no combustion (TN); the observation shows combustion, but the verification shows no combustion (FN); and the observation shows no

combustion, but the verification shows combustion (FP). For TP and TN, which verified fire occurrences correctly, the following formula was used for calculations:

$$P_{accuracy} = n_{correct} / n_{sample} \quad (3)$$

where $n_{correct}$ is the number of correctly verified burning events, n_{sample} is the size of the overall sample, and $P_{accuracy}$ stands for the accuracy of this verification.

We applied receiver operating characteristic (ROC) analysis to express the forecasting ability of each model based on all factors. Plotting ROC curves using real and spurious signals (sensitivity and 1-specificity), the predictive ability of the model can be gauged by quantifying the area under the ROC curve line [37,38]. An AUC value lower than 0.60 is considered to be a poor forecaster, a value between 0.60 and 0.75 may be good for forecasting, and a value above 0.75 indicates that the model has a good predictive capability [39].

In a binary selection model, the judgment threshold (cut-off) is critical for evaluating the prediction probability of the model. In the past, many studies have used a system default value of 0.5 as the evaluation standard, although some recent studies have warned that this may cause large deviations. To overcome this problem, the cut-off point was determined using the Yueden criterion [29], which is a threshold for estimating a model's forecasting accuracy and allows probability to be classified. This method has been used for research into the forecasting of forest fires [40–42]; for example, when the forecasted likelihood of the model is higher than the cut-off point, a field fire is presumed to be burning at that location, with the opposite being true if the forecasted probability is less than the cut-off value. The calculation of Yueden criteria is based on the specificity and sensitivity. Precision rate, recall rate, and the positive (LR+) and negative (LR-) likelihood ratio for validation datasets in all three models were computed according to these cut-off values. The precision rate describes the ratio of samples that burned compared to those verified to have burned, and the recall rate describes the ratio of correctly verified fire points to the number of fire observations. As such, higher values for both factors denote better model performance. LR+ and LR- are two additional parameters used to evaluate the quality of the model; here, large values for LR+ denote better model performance, although the opposite is true for LR-. The formulae used for calculations are as follows:

$$Cut - off = sensitivity + specificity - 1 \quad (4)$$

$$\begin{aligned} P_{precision} &= n_{TP} / (n_{TP} + n_{FP}) \\ P_{recall} &= n_{TP} / (n_{TP} + n_{FN}) \\ P_{positive} &= n_{TP} / n_{FP} \\ P_{negative} &= n_{FN} / n_{TN} \end{aligned} \quad (5)$$

where *cut-off* refers to the value of the cut-off point; $P_{precision}$, P_{recall} , $P_{positive}$, and $P_{negative}$ represent the precision rate, recall rate, positive, and negative of this modeling, respectively; and n_{TP} , n_{FP} , n_{FN} , and n_{TN} represent the numbers of cases for TP, FP, FN, and TN, respectively.

3. Results

3.1. Frequency of Fire Events

A total of 31,944 crop residue open burning fire points occurred in Northeastern China from 16 March to 20 April between 2018 and 2021. The number and spatial distribution of fire points varied each year, with the Himawari-8 observations recording 1202, 483, 5777 and 24,482 events in 2018, 2019, 2020 and 2021, respectively. Figure 2 shows that under the influence of anthropogenic management and control policy, the fires in 2018 mainly occurred in the western part of Heilongjiang Province and on the border with Jilin Province. In 2019, fire events were reduced, and most fires occurred in the northwestern parts of Heilongjiang Province, and the northeastern parts of Heilongjiang and Jilin Provinces. The most occurred around central Jilin Province, and the northwestern and southeastern parts

of Heilongjiang Province in 2020. In addition to the northwestern of Heilongjiang Province, the western and northern of Inner Mongolia, the number of crop residue fires increased significantly in 2021.

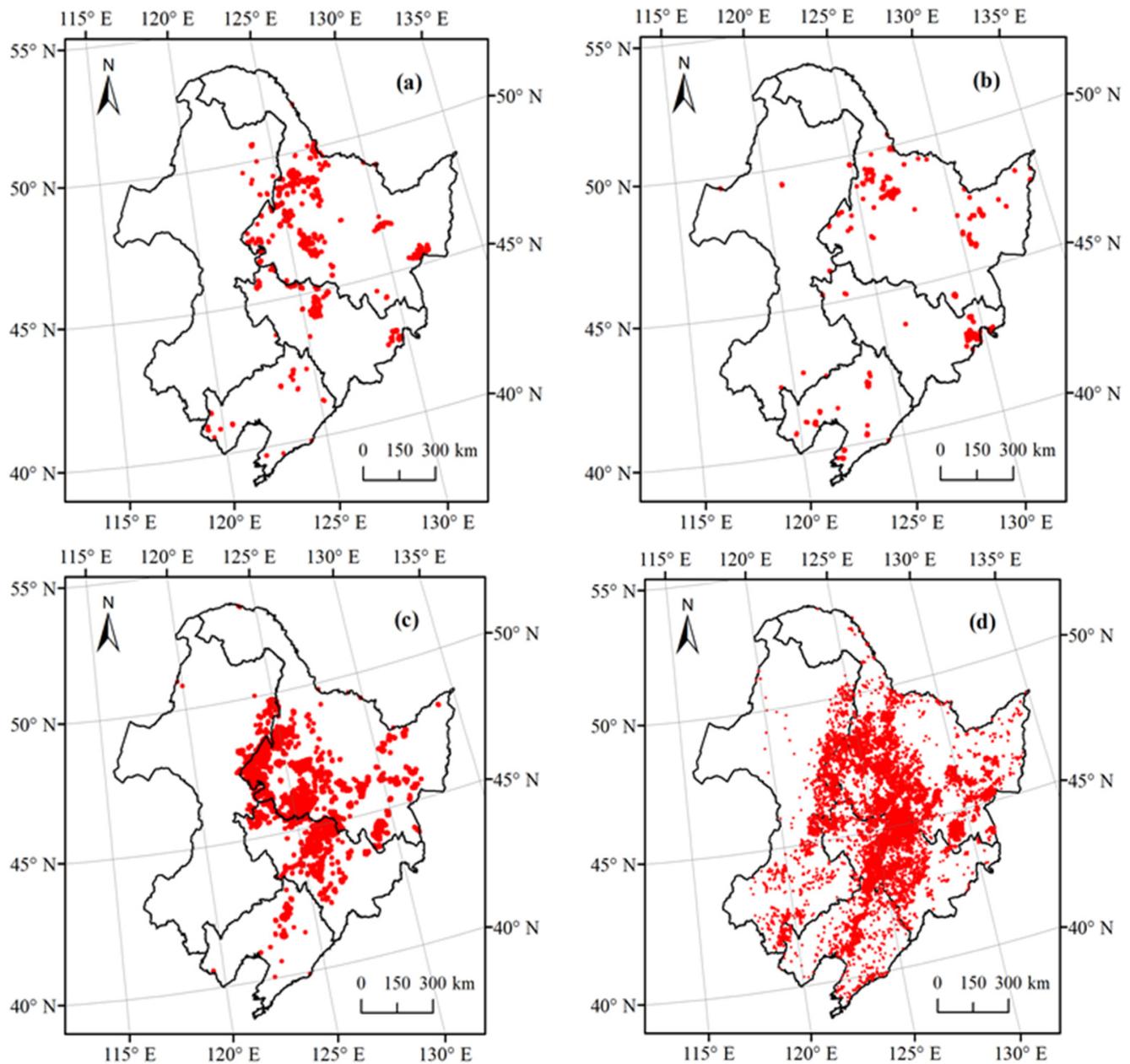


Figure 2. Fire points for open crop burning observed by Himawari-8 between 16 March and 20 April each year in 2018 (a), 2019 (b), 2020 (c) and 2021 (d).

3.2. Comparison of Verification Accuracy between Models

While 48,964 data points were present in the verification set, only 24,482 crop residue open burning fire points were observed by the Himawari-8 Monitor. As shown in Table 2, LR, BPNN, and DT models verified 17,728, 28,072, and 20,431 crop residue open burning fire points, respectively, meaning that the BPNN and DT techniques provided closer matches to the Himawari-8 observations. The verification accuracy of each method was 58.16, 68.59, and 79.59%, respectively. The DT and BPNN models showed a higher forecasting accuracy than LR, reflecting the existence of nonlinear trends in the data. These results reveal that LR

has a lower forecasting effect than DT and BPNN, and that DT is the most suitable method for forecasting crop residue open burning.

Table 2. Results for verification of fire points over Northeastern China in 2021 by the LR, BPNN, and DT method.

Methods	Sort	Verified Fire Points	TP	TN	FN	FP
LR	Samples	17,228	10,612	17,866	13,870	6616
	Proportion (%)	35.10	21.67	36.49	28.33	13.51
BPNN	Samples	28,072	18,586	14,996	5896	9486
	Proportion (%)	57.30	37.96	30.63	12.04	19.37
DT	Samples	20,431	17,461	21,512	7021	2970
	Proportion (%)	41.70	35.66	43.93	14.34	6.07

TP—both verification and observations indicate combustion; TN—both verification and observations indicate no combustion; FN—observations indicate combustion, but verification does not; and FP—verification indicates combustion, but observations do not.

3.3. Causes and Analysis of False Fire Results

The spatial distribution of Himawari-8 observations and the three methods used to verify crop residue fire points from 16 March to 20 April of 2021 are shown in Figure 3. The fire verification locations of the DT method were generally consistent with the observed locations, the BPNN method verified fire locations in the northwestern of Heilongjiang and Jilin Province, the eastern side of Inner Mongolia and Liaoning Province were more than the observed, although the LR and BPNN models did not verify fires correctly. A comparison between verification and the observation data for the DT method showed slight differences in the central northwestern regions of Jilin Province. When the spatial extent of the straw open burning prohibition areas in Jilin Province are considered, it becomes clear that the fire points that had not been forecasted by our models were mainly distributed in the open burning prohibition areas. Thus, as the models will automatically determine that there will be no fire in these areas, this mismatch does not represent a problem for the model forecasting process.

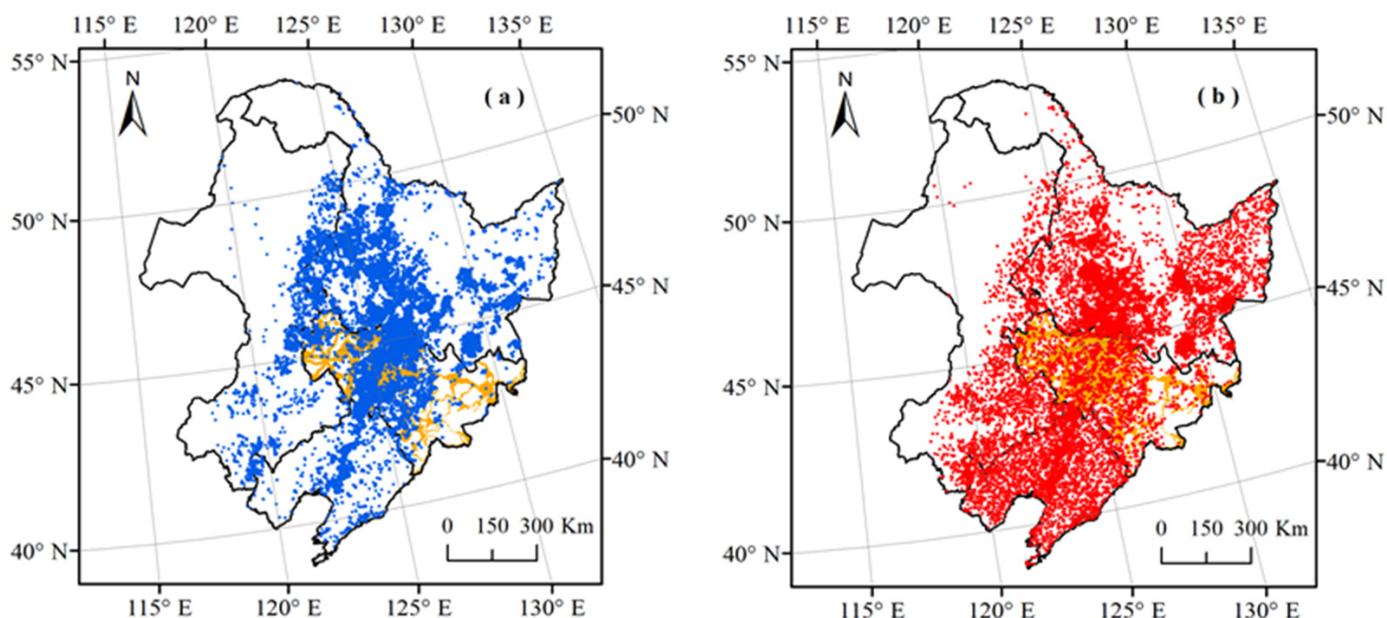


Figure 3. Cont.

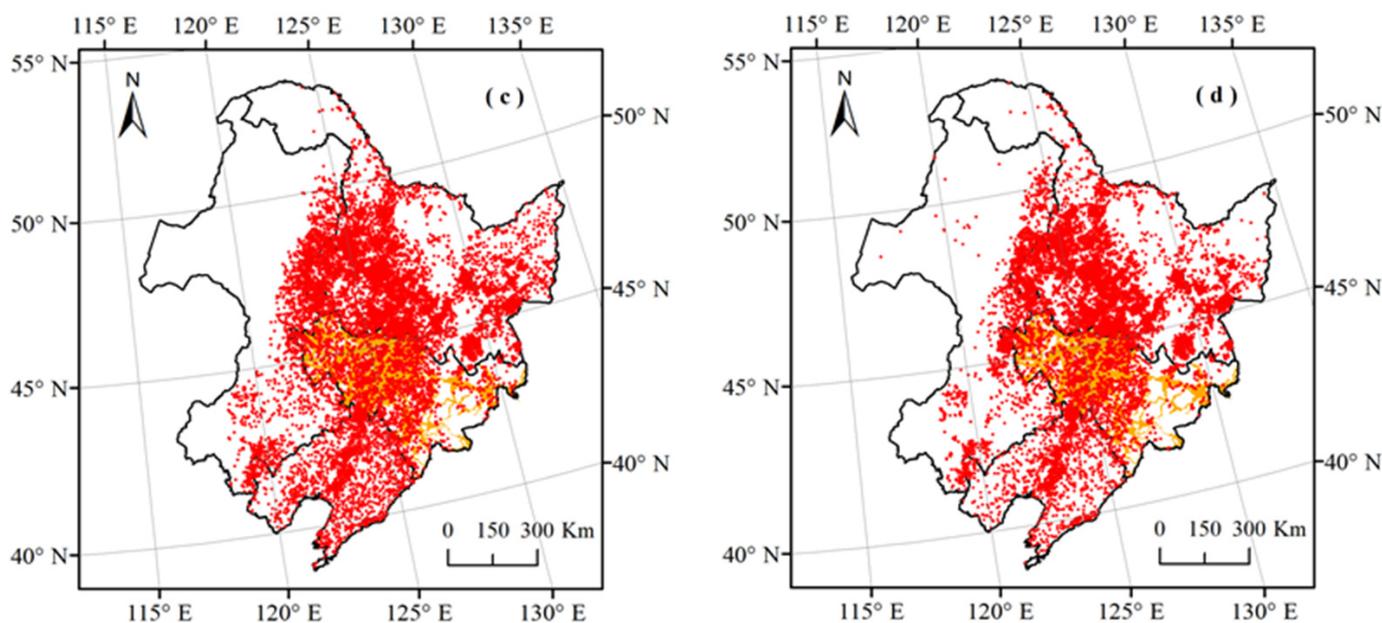


Figure 3. Spatial distribution of observation and verification field burnings in Northeastern China compared from 16 March to 20 April in 2021; (a) shows observed data by the Himawari-8 satellite (blue points); (b–d) present verified data produced by LR, BPNN and DT methods, respectively (red points), as noted by the inset abbreviations. Orange shading represents the straw open burning prohibition areas within Jilin Province.

4. Discussion

4.1. Forecasting Improvement and Model Evaluation

Our previously published research on forecasting of crop residue open burning in Northeastern China using the BPNN method reported a verification accuracy of 60% [16]. In this study the verified general accuracy was 68.59%, and the proportions for TP and TN were 37.96 and 30.63%, respectively. This result marks an improvement in the forecasting ability from our previous work. Nonetheless, the sample size used for training (14,924) in this study was much higher than the previous study (248 for training). This confirms our previous conjecture, whereby an increase in sample size can improve the forecasting ability of BPNN [13,14]. This comparison also suggests that a larger training data set means that the neural network will have a better learning performance, at least within a certain range of sample sizes. This statement is reasonable and agrees with previously published work [13].

A series of ROC curves were constructed to assess the forecasting capabilities of LR, BPNN, and DT models, and to calculate AUC values to test the modeling suitability of each. The ROC curve for each model is shown in Figure 4a, where it can be seen that the ROC curve for DT is positioned above those for LR and BPNN models, implying that DT is the best forecasting tool. This is consistent with that presented in Table 2, where the verification accuracy of DT (79.59%) was higher than that for LR (58.16%) and BPNN (68.59%) using the validation data set. Previous studies assessing forest fires have reached similar conclusions concerning the strong predictive capacity of the DT technique [31,43].

Additional evaluation indicators were calculated for the LR, BPNN, and DT models, which are shown in Table 3. According to these calculations, the sensitivity and specificity values of the DT model were greater than 60 and 80%, respectively, with this result illustrating that the model is stable. In addition, the AUC value of DT was higher than 0.75, which also indicates a high goodness of fit for the model. BPNN and DT models had good recall rates and low probabilities of reported fire points being missing, which is highly significant for anthropogenic management and control of crop residue open burning. Furthermore, the low precision rate of the DT model (44.8%) suggests that most of the correctly verified

fire occurrences were TN (i.e., both verification and observation data indicated no burning). The precision rate of the BPNN model was 55.34%, which was the best result among all three models, and it is notable that DT was much more informative than BPNN and LR, with the highest LR+ at 4.30 and the lowest LR- value at 0.39.

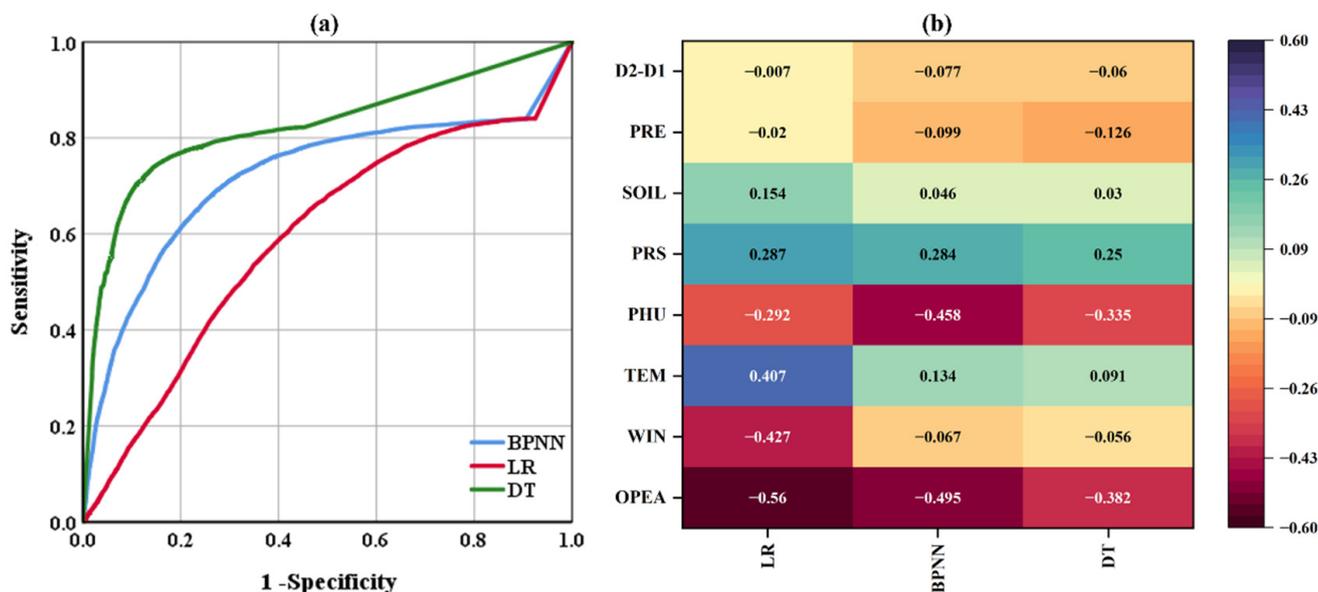


Figure 4. ROC curves (a) and heatmap of impact factor correlations (b) for LR, BPNN, and DT models in this study.

Table 3. Results of verification accuracy and fit superiority of LR, BPNN, and DT models for all variables across Northeastern China in 2021.

Model Classification	AUC	Sensitivity	Specificity	LR+	LR-	Cut-Off Value	Accuracy	Precision	Recall
LR	0.72	69.40	63.38	1.90	0.48	0.33	58.16	37.26	43.35
BPNN	0.59	53.53	73.45	2.02	0.63	0.27	70.82	55.34	75.92
DT	0.82	67.24	84.37	4.30	0.39	0.52	79.59	44.8	71.32

AUC—area under the curve; LR+—positive likelihood ratio; LR—negative likelihood ratio.

4.2. Correlation of Factors Affecting Combustion

The correlation coefficient is a statistic measure commonly applied to show the relationship between the dependent variable and the independent variables. Its squared value is frequently used as a measure of the goodness-of-fit of a model [44]. There are a range of correlation coefficients that are based on different statistical hypotheses; here, we used the Pearson’s correlation coefficient to quantify the relationship between input variables and fire point forecasting factors for the LR, BPNN, and DT methods. The Pearson correlation coefficient is named after Karl Pearson, who was a major contributor to the development of statistics. The Pearson correlation coefficient values range between -1 and 1 ; values greater than 0 show that the dependent variable is generally positively correlated with the independent variable, whereas values less than 0 show that the dependent variable is negatively correlated with the independent variable. Estimating which factors influence field burning occurrence is essential for exploring the spatial distribution of agricultural fires and their causes. Eight different input variables have been compared with verification results to assess which factors have the greatest influence on crop residue open burning in Northeastern China. The relative performance of the three ML methods used in this study were statistically characterized by using the Pearson correlation coefficient, with results shown in Table 4, while a heat map is drawn in Figure 4b.

Table 4. Relationship of input data and agricultural fire forecasting factors in all three models.

Methods	Sort	Correlation of the Input Variables							
		WIN	PRE	TEM	PRS	PHU	SOIL	D2-D1	OPEA
LR	R	−0.427	−0.02	0.407	0.287	−0.292	0.154	−0.007	−0.56
	Sig.	0	0.264	0	0	0	0	0.108	0
BPNN	R	−0.067	−0.099	0.134	0.284	−0.458	0.046	−0.077	−0.495
	Sig.	0	0	0	0	0	0	0	0
DT	R	−0.056	−0.126	0.091	0.25	−0.335	0.03	−0.06	−0.382
	Sig.	0	0	0	0	0	0	0	0

R—the Pearson correlation coefficient value; Sig.—significance, where $p < 0.05$ is significant, $p < 0.01$ is extremely significant, and $p > 0.05$ is not significant.

The calculated values of the Pearson’s correlation coefficient show that TEM, PRS, and SOIL generally showed positive correlations with fire points, but WIN, OPEA, PHU, PRE, and D2-D1 showed negative correlations instead. The largest positive effect on combustion was caused by PRS, and the biggest rate of negative change on open burning was caused by OPEA. For DT and BPNN methods, all input variables showed extremely significant correlations with crop residue open burning. Both methods show a very similar performance in evaluation of rankings for verification fire points, with OPEA ranked first for all methods, followed by PHU, PRS, TEM/PRE, D2-D1, WIN and SOIL. The most descriptive features were OPEA, PHU, and PRE, it is straightforward to understand these negatively correlate with crop residue open burning events, Although the reason for a strong correlation between TEM or PRS and such fires is unclear, we suspect that a change in air temperature influences non-linear feedback between pressure and relative humidity, and directly affects many factors that in turn influence the likelihood of fire occurrence. These findings are consistent with those of previous studies [13,45,46].

4.3. Advantages and Limitations of ML Methods for Fire Forecasting

This study has successfully used ML methods to correlate partial field fire points in Northeastern China between 2018 and 2021 with natural and anthropogenic data. We have shown that ML models can forecast crop residue open burning events at the regional level and are capable of forecasting fires and farmland burning more efficiently than field investigation. A well-trained and verified DT or BPNN can be employed to forecast the occurrence of field fires, which can eliminate considerable investment of time and money in practical activities. In this study, we developed ML models that considered anthropogenic management and control factors, and which were built on a previous BPNN model. The results of our models provide important new ideas for efficient forecasting of agricultural fires, and offer an in-depth and detailed analysis of which factors are most influential in optimizing ML design and operation.

Despite this, ML techniques have limitations. For example, regardless of which approach is taken, substantial effort is required to select and optimize the model parameters, with parameter selection being mostly reliant on experience and lacking clear theoretical guidelines. Although ML is highly capable of performing data analysis, fitting, and making predictions, training effectiveness depends heavily on the accuracy of the input parameters. As such, data must be filtered to remove outliers and considerable improvements are required for standardization. In addition, merely the size of the training set determines the learning effect, whereby small samples limit the forecasting ability of the algorithm, thus leading to poor generalization [47]. In the training set used for BPNN construction, numerical fluctuations between repeated results under the same combination of design parameter settings were inevitable, such that the forecasted results also showed a range of fluctuations [48,49].

4.4. Prospects for the Future of Model Application

The a priori identification of crop residue open burning is one of the most important technical ways to control air pollution and manage anthropogenic changes. BPNN and DT models can immediately forecast open burning of crop residue by using meteorological data, which can be obtained from existing air environment quality models based in any country worldwide, and data for other natural or anthropogenic factors can be obtained from other sources. The assessment of crop residue fire occurrences at this scale can provide appropriate preventive measures for local government regulators and other decision makers to deal with field fires; in particular, by designing agricultural waste management strategies that are tailored to different conditions in different regions. Thus, the methods proposed here can provide rapid, science-based, and technical support for a government to control open burning of crop residue in the future. The results of forecasting emissions from fires may be applied to atmospheric environment models, after converting between scales, which may then provide improved data for these atmospheric models to improve their forecasting ability. Effective use of such converted results may also significantly reduce the numbers of acquisition stations needed for measurement and data acquisition.

Simultaneously, forecasting crop residue open burning events is critical for prevention and management practices in Northeastern China, such that this topic is worthy of continued research. Subsequent work could concentrate on predicting changes across various spatial and temporal dimensions. Deep learning as a kind of machine learning, it has stronger learning ability and adaptability, and the image recognition field performs particularly well. Although there have been studies using CNN methods to forecast forest fires on a regional [50,51] and global scale [52], there is still a gap in the forecasting of crop fires. In the future study, we would consider this method or a combination of multiple machine learning to forecast the fire points of crop residues.

5. Conclusions

In this study, we compared the verification accuracy of logistic regression (LR), back-propagation neural network (BPNN), and decision tree (DT) models for fire point activities of crop residue open burning. The optimal method was determined by verifying field fires with Himawari-8 observed daily fire pixel counts from 2021 in Northeastern China. The verification results indicated that BPNN and DT models produced higher verification accuracies when compared to those of previous studies. Indeed, verification accuracies for BPNN and DT models both exceeded 70%, although the DT model had a statistically significant higher value of 79.59%. By analyzing the effects of combustion factors, this study shows that the existence of open burning prohibition areas, average daily relative humidity and average daily air pressure were the most significant factors that affected forecasting accuracy. By contrast, the soil moisture content on the day and change in soil moisture content over a period of 24 h contributed to a lesser degree to the forecasting of crop residue fires at the local scale. This study improves our ability to forecast agricultural field fires throughout Northeastern China, and provides a scientific framework for the regional control of crop residue open burning. The results are not only available to design agricultural waste management strategies suitable for various situations in different regions, but also to predict crop emissions according to the management strategies in areas where they are not used. They could even provide scientific and technical support for air quality prediction.

Taken together, machine learning can be used to forecast crop residue open burning on a regional scale for areas that are experiencing anthropogenic management and control. Implementation of the DT method appears to be the most accurate and reliable investigation tool. The SPSS modeler tool that we chose this time is simple and intuitional to operate, but its operation speed is relatively slow. Combined with the powerful recognition capabilities of deep learning in the image field, subsequent work can be explored based on this advantage to forecast changes in different spatial and temporal dimensions.

Author Contributions: B.B.: methodology, software, validation, formal analysis, data curation, writing—original draft preparation, writing—review and editing, visualization. H.Z.: conceptualization, validation, visualization, supervision. S.Z.: validation, visualization. X.L.: data curation. X.Z.: validation, visualization. A.X.: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work is financially supported by the National Natural Science Foundation of China (No. 42101414) and the National Natural Science Foundation of Jilin Province (No. 20200201214JC).

Data Availability Statement: In this paper, some data is available, and the rest is unavailable: The fire and soil moisture data that support this study were obtained from JAXA Himawari Monitor P-Tree System and European Space Agency by permission/license. Upon reasonable request and with the permission of JAXA Himawari Monitor P-Tree System and European Space Agency, the data will be shared with the respective author. The meteorological and anthropogenic data supporting this study are not available for public sharing for moral or privacy reasons, and if appropriate, can be made available upon legitimate request to the corresponding author.

Acknowledgments: We acknowledge the JAXA Himawari Monitor P-Tree System, European Space Agency, China Meteorological Data Network and People's Government of Jilin Province freely sharing the wildfire, soil moisture content, meteorological and anthropogenic management and control policy data.

Conflicts of Interest: The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Yin, S.; Wang, X.F.; Xiao, Y.; Tain, H.; Zhong, G.S.; Sun, Z.Y. Study on spatial distribution of crop residue burning and PM2.5 change in China. *Environ. Pollut.* **2017**, *220*, 204–221. [[CrossRef](#)] [[PubMed](#)]
2. Zhao, H.M.; Zhang, X.L.; Zhang, S.C.; Chen, W.W.; Tong, D.Q.; Xiu, A.J. Effects of agricultural biomass burning on regional haze in China: A review. *Atmosphere* **2017**, *8*, 88. [[CrossRef](#)]
3. Wang, G.S.; Gao, B. Fire location model based on adaptive learning rate BP Neural Network. *Comput. Syst. Appl.* **2019**, *28*, 250–254, (In Chinese with English Abstract). [[CrossRef](#)]
4. Satir, O.; Berberoglu, S.; Donmez, C. Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem. *Geomat. Nat. Haz. Risk.* **2016**, *7*, 1645–1658. [[CrossRef](#)]
5. Barboza, C.E.; Turpo, E.Y.; de Almeida, C.M.; Salas, L.R.; Rojas, N.B.; Silva, L.; Jhonsy, O.; Barrena, M.A.; Oliva, M.; Espinoza-Villar, R. Monitoring wildfires in the Northeastern Peruvian Amazon using Landsat-8 and Sentinel-2 imagery in the GEE Platform. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 564. [[CrossRef](#)]
6. Zhang, G.L.; Wang, M.; Liu, K. Forest fire susceptibility modeling using a convolutional neural network for yunnan province of China. *Int. J. Disast. Risk Sci.* **2019**, *10*, 386–403. [[CrossRef](#)]
7. Michael, Y.; Helman, D.; Glickman, O.; Gabay, D.; Brenner, S.; Lensky, I.M. Forecasting fire risk with machine learning and dynamic information derived from satellite vegetation index time-series. *Sci. Total Environ.* **2021**, *764*, 142844. [[CrossRef](#)]
8. Li, L.M.; Song, W.G.; Ma, J.; Kohyu, S. Artificial neural network approach for modeling the impact of population density and weather parameters on forest fire risk. *Int. J. Wildland Fire* **2009**, *18*, 640–647. [[CrossRef](#)]
9. Alkabbani, H.; Ramadan, A.; Zhu, Q.Q.; Elkamel, A. An improved air quality index machine learning-based forecasting with multivariate data imputation approach. *Atmosphere* **2022**, *13*, 1144. [[CrossRef](#)]
10. Jain, P.; Coogan, S.C.P.; Subramanian, S.G.; Crowley, M.; Taylor, S.; Flannigan, M.D. A review of machine learning applications in wildfire science and management. *Environ. Rev.* **2020**, *28*, 478–505. [[CrossRef](#)]
11. Liu, F.; Xu, L.; Ma, X.X. Development of BP neural network and its application in chemistry and chemical engineering. *Chem. Ind. Eng. Prog.* **2019**, *38*, 2559–2573, (In Chinese with English Abstract). [[CrossRef](#)]
12. Di Giuseppe, F.; Rémy, S.; Pappenberger, F.; Wetterhall, F. Improving Forecasts of Biomass Burning Emissions with the Fire Weather Index. *J. Appl. Meteorol. Clim.* **2017**, *56*, 2789–2799. [[CrossRef](#)]
13. Feng, X.; Fu, T.M.; Cao, H.S.; Tian, H.; Fan, Q.; Chen, X.Y. Neural network predictions of pollutant emissions from open burning of crop residues: Application to air quality forecasts in southern China. *Atmos. Environ.* **2019**, *204*, 22–31. [[CrossRef](#)]
14. Bai, B.; Zhao, H.M.; Zhang, S.M.; Zhang, X.L.; Yang, G.Y. Forecasting of agricultural straw burning in the Northeastern China based on neural network. *China Environ. Sci.* **2020**, *40*, 5205–5212, (In Chinese with English Abstract). [[CrossRef](#)]
15. Wen, X.; Chen, W.W.; Chen, B.; Yang, C.J.; Tu, G.; Cheng, T.H. Does the prohibition on open burning of straw mitigate air pollution? An empirical study in Jilin Province of China in the post-harvest season. *J. Environ. Manag.* **2020**, *264*, 110451. [[CrossRef](#)]
16. Bai, B.; Zhao, H.M.; Zhang, S.M.; Zhang, X.L.; Du, Y. Can neural networks forecast open field burning of crop residue in regions with anthropogenic management and control? A case study in Northeastern China. *Remote Sens.* **2021**, *13*, 3988. [[CrossRef](#)]

17. Li, H.; Zhang, H.Y.; Li, Q.X.; Zhao, J.J.; Guo, X.Y.; Ying, H.; Deng, G.R.; Wu, R.H.; Wang, S.L. Vegetation productivity dynamics in response to climate change and human activities under different topography and land cover in Northeast China. *Remote Sens.* **2021**, *13*, 975. [CrossRef]
18. Xu, X.Z.; Xu, Y.; Chen, S.C.; Xu, S.G.; Zhang, H.W. Soil loss and conservation in the black soil region of Northeast China: A retrospective study. *Environ. Sci. Policy* **2010**, *13*, 793–800. [CrossRef]
19. Zhang, H.; Zhou, G.S.; Liu, D.L.; Wang, B.; Xiao, D.P.; He, L. Climate-associated rice yield change in the Northeast China Plain: A simulation analysis based on CMIP5 multi-model ensemble projection. *Sci. Total Environ.* **2019**, *666*, 126–138. [CrossRef]
20. Wang, J.W.; Tang, H.; Wang, J.F. Comprehensive utilization status and development analysis of crop straw resource in Northeast China. *Trans. Chin. Soc. Agric. Mach.* **2017**, *48*, 1–21, (In Chinese with English Abstract). [CrossRef]
21. Zhang, J.Y.; Yang, X.H.; Tu, X.M.; Ning, K.X.; Luan, X.C. Spatio-temporal change of straw burning fire points in field of China from 2014 to 2018. *Trans. Chin. Soc. Agric. Eng.* **2019**, *35*, 191–199, (In Chinese with English Abstract). [CrossRef]
22. de Vasconcelos, M.J.P.; Silva, S.; Tome, M.; Alvim, M.; Pereira, J.M.C. Spatial prediction of fire ignition probabilities: Comparing logistic regression and neural networks. *Photogramm. Eng. Rem. S* **2001**, *67*, 73–81.
23. JAXA Himawari Monitor P-Tree System. Available online: <https://www.eorc.jaxa.jp/ptree/> (accessed on 7 August 2020).
24. Bessho, K.; Date, K.; Hayashi, M.; Ikeda, A.; Imai, T.; Inoue, H.; Kumagai, Y.; Miyakawa, T.; Murata, H.; Ohno, T.; et al. An introduction to Himawari-8/9-Japan's new-generation geostationary meteorological satellites. *J. Meteorol. Soc. Jpn.* **2016**, *94*, 151–183. [CrossRef]
25. China Meteorological Data Network Home Page. Available online: <https://data.cma.cn/> (accessed on 20 June 2020).
26. European Space Agency, Climate Change Initiative Soil Moisture Project Home Page. Available online: <http://www.esa-soilmoisturecci.org/> (accessed on 8 December 2020).
27. Dorigo, W.; Wagner, W.; Albergel, C.; Albrecht, F.; Balsamo, G.; Brocca, L.; Chung, D.; Ertl, M.; Forkel, M.; Gruber, A.; et al. ESA CCI soil moisture for improved earth system understanding: State-of-the art and future directions. *Remote Sens. Environ.* **2017**, *203*, 185–215. [CrossRef]
28. Jilin Province Government, Work Program for Straw Open Burning Prohibition in Autumn and Winter of 2018. Available online: http://sthjt.jl.gov.cn/xxfb/tzgg/201810/t20181024_5196951.html/ (accessed on 10 July 2020).
29. Vega-Garcia, C.; Woodard, P.M.; Titus, S.J.; Adamowicz, W.L.; Lee, B.S. A logit model for predicting the daily occurrence of human caused forest-fires. *Int. J. Wildland Fire* **1995**, *5*, 101–111. [CrossRef]
30. Martinez, J.; Vega-Garcia, C.; Chuvieco, E. Human-caused wildfire risk rating for prevention planning in Spain. *J. Environ. Manag.* **2009**, *90*, 1241–1252. [CrossRef]
31. Oliveira, S.; Oehler, F.; San-Miguel-Ayanz, J.; Camia, A.; Pereira, J.M.C. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *For. Ecol. Manag.* **2012**, *275*, 117–129. [CrossRef]
32. Rodrigues, M.; de la Riva, J.; Fotheringham, S. Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression. *Appl. Geogr.* **2014**, *48*, 52–63. [CrossRef]
33. Liu, W.B.; Wang, Z.D.; Liu, X.H.; Zeng, N.Y.; Liu, Y.R.; Alsaadi, F.E. A survey of deep neural network architectures and their applications. *Neurocomputing* **2017**, *234*, 11–26. [CrossRef]
34. Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2792. [CrossRef]
35. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
36. Groemping, U. Variable importance assessment in regression: Linear regression versus random forest. *Am. Stat.* **2009**, *63*, 308–319. [CrossRef]
37. Jimenez-Valverde, A. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Glob. Ecol. Biogeogr.* **2012**, *21*, 498–507. [CrossRef]
38. Chang, Y.; Zhu, Z.L.; Bu, R.C.; Chen, H.W.; Feng, Y.T.; Li, Y.H.; Hu, Y.M.; Wang, Z.C. Predicting fire occurrence patterns with logistic regression in Heilongjiang Province, China. *Landsc. Ecol.* **2013**, *28*, 1989–2004. [CrossRef]
39. Shukla, A.; Kumar, S.; Singh, H. MLP-ANN-Based execution time prediction model and assessment of input parameters through structural modeling. *Proc. Natl. Acad. Sci. India Sect. A Phys. Sci.* **2021**, *91*, 577–585. [CrossRef]
40. Littell, J.S.; McKenzie, D.; Peterson, D.L.; Westerling, A.L. Climate and wildfire area burned in western U.S. ecoprovinces, 1916–2003. *Ecol. Appl.* **2009**, *19*, 1003–1021. [CrossRef]
41. Guo, F.T.; Wang, G.Y.; Su, Z.W.; Liang, H.L.; Wang, W.H.; Lin, F.F.; Liu, A.Q. What drives forest fire in Fujian, China? Evidence from logistic regression and Random Forests. *Int. J. Wildland Fire* **2016**, *25*, 505–519. [CrossRef]
42. Chen, D. Prediction of forest fire occurrence in Daxing'an Mountains based on logistic regression model. *For. Resour. Manag.* **2019**, *1*, 116–122, (In Chinese with English Abstract). [CrossRef]
43. Bui, D.T.; Bui, Q.T.; Nguyen, Q.P.; Pradhan, B.; Nampak, H.; Trinh, P.T. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agr. For. Meteorol.* **2017**, *233*, 32–44. [CrossRef]
44. Bolboaca, S.D.; Jäntschi, L. Pearson versus Spearman, Kendall's Tau correlation analysis on structure-activity relationships of biologic active compounds. *Leonardo J. Sci.* **2006**, *5*, 179–200.
45. Vasilakos, C.; Kalabokidis, K.; Hatzopoulos, J.; Matsinos, I. Identifying wildland fire ignition factors through sensitivity analysis of a neural network. *Nat. Hazards* **2009**, *50*, 125–143. [CrossRef]

46. Maeda, E.E.; Formaggio, A.R.; Shimabukuro, Y.E.; Arcoverde, G.F.B.; Hansen, M.C. Predicting forest fire in the Brazilian Amazon using MODIS imagery and artificial neural networks. *Int. J. Appl. Earth Obs.* **2009**, *11*, 265–272. [[CrossRef](#)]
47. Guo, C.Q.; Cui, Y.L. Utilizing artificial neural network to simulate and predict the hydraulic performance of free water surface constructed wetlands. *J. Environ. Manag.* **2021**, *305*, 114334. [[CrossRef](#)]
48. Akrotos, C.S.; Papaspyros, J.N.E.; Tsihrintzis, V.A. Total nitrogen and ammonia removal prediction in horizontal subsurface flow constructed wetlands: Use of artificial neural networks and development of a design equation. *Bioresour. Technol.* **2009**, *100*, 586–596. [[CrossRef](#)]
49. Witek-Krowiak, A.; Chojnacka, K.; Podstawczyk, D.; Dawiec, A.; Pokomeda, K. Application of response surface methodology and artificial neural network methods in modelling and optimization of biosorption process. *Bioresour. Technol.* **2014**, *160*, 150–160. [[CrossRef](#)]
50. Cheng, S.B.; Jin, Y.F.; Harrison, S.; 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](#)]
51. Cheng, S.B.; Prentice, I.C.; Huang, Y.H.; Jin, Y.F.; Guo, Y.K.; Arcucci, R. Data-driven surrogate model with latent data assimilation: Application to wildfire forecasting. *J. Comput. Phys.* **2022**, *464*, 111302. [[CrossRef](#)]
52. Zhu, Q.; Li, F.; Riley, W.J.; Xu, L.; Zhao, L.; Yuan, K.X.J.; Wu, H.Y.; Gong, J.Y.; Randerson, J. Building a machine learning surrogate model for wildfire activities within a global Earth system model. *Geosci. Model.* **2022**, *15*, 1899–1911. [[CrossRef](#)]