Article

Modeling Forest Fire Spread Using Machine Learning-Based Cellular Automata in a GIS Environment

Yiqing Xu 1, Dianjing Li 2, Hao Ma 1, Rong Lin 2 and Fuquan Zhang 2,*

1 Industrial Software Engineering Technology Research and Development Center of Jiangsu Education Department, Nanjing Vocational University of Industry Technology, Nanjing 210023, China
2 College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China
* Correspondence: zfq@njfu.edu.cn

Abstract: The quantitative simulation of forest fire spread is of great significance for designing rapid risk management approaches and implementing effective fire fighting strategies. A cellular automaton (CA) is well suited to the dynamic simulation of the spatiotemporal evolution of complex systems, and it is therefore used to model the complex process of forest fire spread. However, the process of forest fire spread is linked with a variety of mutually influencing factors, which are too complex to analyze using conventional approaches. Here, we propose a new method for modeling fire spread, namely LSSVM-CA, in which least squares support vector machines (LSSVM) is combined with a three-dimensional forest fire CA framework. In this approach, the effects of adjacent wind on the law of fire spread are considered and analyzed. The LSSVM is utilized to derive the complex state transformation rules for fire spread by training with a dataset based on actual local data. To validate the proposed model, the forest fire spread area simulated by LSSVM-CA and the actual extracted forest fire spread area were subjected to cross-comparison. The results show that LSSVM-CA performs well in simulating the spread of forest fire and determining the probability of forest fire.

Keywords: cellular automaton (CA); machine learning; forest fire simulating; least squares support vector machines (LSSVM); three-dimensional

1. Introduction

Forest represents a material foundation that is important for a country’s sustainable development, and it is also an important renewable resource in the construction of the economy and ecological environments [1]. Forests provide habitat for wildlife and protect biodiversity [2]. However, forest ecosystems are often subjected not only to natural disturbances but also to manmade damage, including forest diseases, pests, and forest fires [3–5]. Every year, millions of hectares of forests are destroyed by wildfires that seriously threaten human life and property safety [6,7]. When a forest fire is very serious, it will destroy natural ecological species and ecological structure, even threatening human safety and affecting the regional carbon cycle [8–10]. Suppressing wildfires to preserve life and properties has become one of the most critical tasks in preserving natural resources [11,12].

The accurate modeling of forest fire spread facilitates the rational allocation of fire prevention resources and the development of timely and effective fire suppression strategies [13]. In general, forest fire spread models can be classified into three types [14]: remote sensing, physical, and algorithmic.

Remote sensing technology can provide the necessary environmental information required to monitor and forecast fire spreading [15,16]. In remote sensing models, a GIS (geographic information system) is usually used to explore the spatial distribution pattern of forest fires [17]. GIS data have been used to analyze the spatial and temporal distribution characteristics of factors that affect forest fires in different regions [18]. It has been shown that GIS is available for constructing forest fire simulation and rescue systems in formulating corresponding fire prevention measures for local forestry bureaus [19].
Based on the continuous investigations of numerous researchers, many physical models of forest fire behavior based on complex mechanisms have been developed since the concept of forest fire spreading models first appeared in 1946 [20]. FIRETEC and WFDS (Wildland Fire Dynamics Simulator) [21], Physical Modeling of Fire Storms in Russian Federation [22], Meso-NH model in Portugal [23], and FIRESTAR in France [24–26] are examples of widely used physical models. Due to the harmful nature and instability of forest fires, it is difficult to obtain measurement results of parameters required by some models under uncontrolled conditions. Thus, if a physical model relies on large-scale environmental data, its computational cost and complexity would make infeasible for practical use [27].

There are also empirical models and semi-empirical models that have been developed from physical and empirical models, and these are mainly based on the statistical analysis of existing real data events [28–31]. American scholars put forward the Rothermel model in 1972 through field and laboratory experiments combined with the energy conservation law [32]. The Rothermel model has a total of 11 test parameters, which are used to predict the speed and intensity of fire spread. Although it has a complex model structure and requires a large input, it is still widely used model in fire simulation [33]. With the development of model research, empirical models including McArthur model [34] and CFRRS (Canadian Fire Risk Rating System) [35] were proposed and are now widely used. Although the calculation and implementation of this kind of model are simple, a large number of long-term data measurements are required, and the accuracy of these measurements is highly dependent on the data and may be easy affected by measurement factors.

Recently, the application of algorithmic models is becoming increasingly widespread in the field of forest fire prediction [36,37]. For example, the ARMA (autoregressive moving average) model was used to predict and simulate the damage area of forest fires, and it was found that forest fires exhibit an outbreak cycle in China [38]. Using forest fire data in the Daxinganling region from 1990 to 2009 and based on the principle of SVM (support vector machine), a forest fire risk prediction model was derived for local forestry departments to develop fire prevention strategies [39]. Least squares support vector machines is an improved method based on SVM for forest fire prediction using historical forest fire data [40]. Bayesian network has also been used to model and predict the possible causes of forest fires and analyze the interactions between these possible causes [41–43]. Use of a model with a self-learning scheme was shown to have higher accuracy than the model without this scheme [44–47]. In addition, the continuous development of fire spread models [48] since the 1970s has facilitated the emergence and application of fire growth simulation software, such as Prometheus and Farsite [49,50], which are important fire management tools for predicting future fire boundaries and associated fire spread behavior.

Cellular automata are well suited to the dynamic simulation of the spatiotemporal evolution of complex systems and have also been applied to reproduce the evolution of natural phenomena, such as the spread of epidemics [51,52] and in ecological [53–55] and urban growth [56,57] modeling.

In physical models, cellular automata are an idealized physical system model that is discrete, not only in time but also in space and whose physical parameters only take numerical sets.

Cellular automata were first used to develop a suitable model for large-scale parallel computation [58]. Then, they were used in the simulation of forest fires, as they allow modeling of more complicated forest fire spread mechanisms without the need for an in-depth preliminary study of their general evolutionary causes. As a bottom-up modeling method, cellular automata can simulate the spatiotemporal dynamics of spatially complex systems by simply establishing basic transformation rules between neighboring local cells [28,59] based on the ideal heat transfer law between burning and non-burning adjacent cells under the same conditions [60]. Moreover, the model has been improved by extending the conversion rule to different cases such that it can obtain different burning probabilities for top and surface fires [61].
In most studies, optimization of the local state transition rule is one of the key research issues in the simulation of fire spread [62]. The traditional local transition rule for square lattice CA models was updated [63] and then extended for hexagonal dot-matrix CA modeling [64].

Considering a regular grid and constant wind direction may lead to fire spread along the axis of the grid rather than the real direction of the wind, leading to the proposal of a cell-based simulator with an irregular grid [65] to allow the modeled shape of fire spread to be closer to that of the real observed elongated fire shape. In addition, the neighborhood of each unit can be extended beyond the neighboring units such that there are more choices for flame propagation direction [66]. Due to the spatial inhomogeneous combustion process of fuel, a hexagonal model of fuel cell network propagation is proposed to predict inhomogeneous fire behavior [67].

In addition, the local transition rule was developed for the CA modeling approach, which takes into account differences in fire spread variability between the same and different materials [68]. In order to alter the linear frontier for the local transfer rule, the cyclic spread of fire fronts from diagonal units was applied [69]. Later, based on a physical modeling approach, the meta-automata rules from vegetation and flame features were defined [70]. In addition, with the development of CA models, a new model based on satellite images at the macroscale and using multi-scale 3D CA to simulate fires was proposed [71], which emphasizes the ability of CA to solve spatial heterogeneity and other aspects. In another way, the effects of fire suppression strategies, fire burn probabilities, and burn processes or fire physical velocity models were incorporated into the transition rules so that the basic CA model [72–74] could be more practical in modeling the spread of fine-scale mountain fires [62,72,75].

Currently, a new algorithmic model combined with a physical model that uses an extreme learning machine (ELM) instead of traditional transformation rules was proposed, named ELM-CA [59], in which data gathered from fires in western USA were used to validate the performance of the model. However, the extreme learning machine is used to solve matrix pseudo-inverses and is only applicable to single hidden layer neural networks. It is prone to overtraining, which results in degradation of the generalization performance.

Research has shown that forest fire spread is a complex process related to a variety of factors. The factors that mainly influence forest fire spread are the combustible material factor, meteorological factor, and topographic factor, which are too complex to analyze using conventional approaches and require the simulation of parallel processing, heuristics, and intelligent thought processes [59,62,76].

In this paper, we propose a machine learning-based forest fire spread model in which the traditional cellular automata framework is combined with LSSVM (least squares support vector machines). In this modeling approach, the effects of adjacent wind on the law of fire spread are considered and analyzed. The proposed model utilizes LSSVM to derive the non-linear transformation rules for fire burning probabilities for the cellular automata. This simplifies the requirement for a physical interaction process and complex heat transfer law evaluation of the evolutionary rules for fire spread. The forest fire spread area simulated using LSSVM-CA and the actual extracted forest fire spread area are subjected to cross-comparison. The results show that the proposed model performs well in simulating the spread of forest fires and determining the probability of forest fire occurrence.

The rest of this article is structured as follows. Section 2 presents the process of modeling the proposed machine learning-based cellular automata by explaining the basic principle and the methods used for modeling. Section 3 presents data processing of the study area and the simulation results and analysis. Section 4 discusses the findings. Section 5 concludes the paper and present avenues for future work.

2. Model Development

The flowchart of the proposed method is shown in Figure 1. First, we used the AETER GDEM 30 m × 30 m resolution digital elevation data and Landsat8 data to extract the
driving influence factor data and fire point data. With these data, LSSVM was trained to
derive the non-linear rule for the probability of fire burning. Meanwhile, the probability
of the influence of meteorological factors on the adjacent CA raster was analyzed and
combined into the CA model to simulate forest fire spread. In order to validate the
proposed model, the burned area of a historical fire was derived using Sentinel 2 data
and used to conduct cross-comparison between the actual burned area with the simulated
burned area [59,77].

![Figure 1](image1.png)

**Figure 1.** The flowchart of the proposed method: (a) data collection; (b) data pre-processing; (c) modeling; (d) acquisition of simulated and actual combustion data; and (e) model validation.

2.1. Principle of Cellular Automata

The square lattice cell is widely used as the cell of the CA framework, as it can simplify
the computation and significantly reduce computational complexity [59]. Moore-type
criterion is used as the criterion for classifying the neighborhood of a two-dimensional cell.
The square grid cell can be divided into four possible states at different discrete time steps:
cannot be burned, has not been ignited, burning, and has been burned (see Figure 2).

![Figure 2](image2.png)

**Figure 2.** The principle of cellular automata: (a) state transition and (b) example of spread.

In a meta-cellular automaton system, the state of each cell evolves in discrete time steps depending on a finite set of state transfer rules. Figure 2a indicates the types of events that lead to each state transition at the next discrete time step (t + 1); the possible transitions of a cell, \( A(i,j) \), are as follows:
• If the cell cannot be burned due to lack of fuel, then it remains in the same state at the next discrete time step.
• If the cell is burning, its state is updated to burned at the next discrete time step.
• If the cell is burned, then it remains in the same state at the next discrete time step.
• If the cell has not been ignited and at least one its neighboring cells is burning \( A(i \pm 1, j \pm 1) \), and if the state transition likelihood \( P_{i,j}^t \) of the cell \( A(i,j) \) is higher than a random probability threshold \( \epsilon_t \), then its state is updated to burning at the next discrete time step.

The random probability threshold in Equation (1) is calculated using the following formula [59,76]:

\[
\epsilon_t = \beta \times \frac{1}{1 + (-\ln \gamma_t)^\alpha}
\]  

(1)

where \( \alpha \) and \( \beta \) are constant values, \( \gamma \) is a random number between 0 and 1.

2.2. Adjacent Wind Effect

In fire spread modeling, wind is widely recognized as an essential factor affecting the velocity and direction of fire spread. Based on the equation of the influence of wind, the coefficient of wind action \( (K_w) \), and wind velocity \( (V) \) can be expressed as [74]

\[
K_w = e^{0.1783V}
\]  

(2)

Due to the varying impact of wind on the spread of fire, it is necessary to derive the wind effect on the eight surrounding neighbor cells, as shown in Figure 3.

Figure 3. Wind speed decomposition model.

Assuming that the wind direction is counterclockwise rotated to \( \vec{OB} \) with an angle of \( \theta \), the projection of the wind blowing in the direction of \( \vec{OB} \) is \( V \cos \theta \). Similarly, the projection components of the wind direction, \( V \), in any direction can be derived. With this and Equation (2), the wind direction coefficients for the eight directions can be expressed as \( \vec{OA} : e^{0.1783 V \cos(315 - \theta)} \), \( \vec{OB} : e^{0.1783 V \cos \theta} \), \( \vec{OC} : e^{0.1783 V \cos(\theta - 45)} \), \( \vec{OD} : e^{0.1783 V \cos(\theta - 90)} \), \( \vec{OE} : e^{0.1783 V \cos(\theta - 135)} \), \( \vec{OF} : e^{0.1783 V \cos(\theta - 180)} \), \( \vec{OG} : e^{0.1783 V \cos(\theta + 225)} \), \( \vec{OH} : e^{0.1783 V \cos(\theta + 90)} \), respectively [74]. Then, the neighboring wind effect \( \theta(i,j)_t \) can be calculated using Equation (3).

\[
\theta(i,j)_t = \frac{\sum_{i=0}^{n} \sum_{j=0}^{n} K_w \times c(i,j)}{\sum_{i=0}^{n} \sum_{j=0}^{n} K_w}
\]  

(3)

where \( c(i,j) \) indicates the cell is burning or not. If the cell is burning, it is 1. Otherwise, it is 0.
2.3. Igniting Probabilities

The classical application area of LSSVM is classification, so it is often used as a classifier. Here, we consider the change of a cell in cellular automaton from one state to another as a classification problem. When studying the spread of forest fires, the main concern is whether the current cell state can be converted to a burning state; if so, then the cell is a fire point at the next moment and is recorded as 1, otherwise the cell is not a fire point at the next moment and is recorded as 0. Therefore, the transition space of the cell is \{0, 1\}, and the conditions for this transition and non-transition are mainly due to the current factors that can affect the cell’s state. Therefore, the influence factor of the cell itself can be recorded as the sample space such that the determination of forest fire spread using the cellular automata transformation rule can be considered a classification problem. That is, the cell transformation value can be judged according to the conditions.

Given a training set of \(N\) samples, where the \(i\)th input data are \(x_i \in \mathbb{R}^n\) and the corresponding \(i\)th output data are \(y_i \in \{0, 1\}\), the goal of a categorical support vector machine is to construct a classifier of the following form:

\[
\tilde{y}(x) = \text{sign} \left[ \omega^T \varphi(x) + b \right]
\]  
(4)

The least squares support vector classification problem is ultimately the optimization problem of solving Equation (5) as follows

\[
\min_{\omega, e} \frac{1}{2} \omega^T \omega + \gamma \sum_{i=1}^{N} e_i^2
\]
and satisfying the equation constraint

\[
y_i \left[ \omega^T \varphi(x) + b \right] = 1 - e_i, \quad i = 1, 2, \ldots, N.
\]  
(6)

The Lagrangian polynomial for the dual problem of Equation (6) is

\[
L(\omega, b, e, \alpha) = J(\omega, e) - \sum_{i=1}^{N} a_i \{ y_i [\omega^T \varphi(x) + b] - (1 - e_i) \}
\]

where \(a_i\) is the Lagrange multiplier, which can be positive or negative due to the equation constraint. The optimality condition is that the partial derivatives are all 0, i.e., Equation (8).

\[
\begin{bmatrix}
  I & 0 & 0 & -Z^T \\
  0 & 0 & 0 & -y^T \\
  0 & 0 & \lambda I & -I \\
  Z & y & I & 0
\end{bmatrix}
\begin{bmatrix}
  \omega \\
  b \\
  e \\
  \alpha
\end{bmatrix}
= \begin{bmatrix}
  0 \\
  0 \\
  0 \\
  \gamma \| \tilde{1} \|
\end{bmatrix}
\]
(8)

In order to eliminate \(e\) and \(\omega\), the Mercer condition will be used here, i.e., Equation (9).

\[
\Omega_{kj} = y_k y_j \varphi(x_k)^T \varphi(x_j) = y_k y_j K(x_k, x_j), \quad k, j = 1, 2, \ldots, N
\]

(9)

The system of equations obtained is related to \(b, \alpha\). Using the above conditions the system of equations is transformed into Equation (10).

\[
\begin{bmatrix}
  0 & -y^T \\
  y & \Omega + \gamma^{-1} I
\end{bmatrix}
\begin{bmatrix}
  b \\
  \alpha
\end{bmatrix}
= \begin{bmatrix}
  0 \\
  \| \tilde{1} \|
\end{bmatrix}
\]
(10)

Assume \(A = \Omega + \gamma^{-1} I\). Where \(A\) is a symmetric semi-positive definite matrix, therefore its inverse matrix \(A^{-1}\) exists. Ultimately, the LSSVM classifier can be expressed as Equation (11).

\[
g(x) = \text{sign} \left[ \sum_{i=1}^{N} a_i y_i K(x, x_i) \right] + b
\]
(11)
Here, the classification decision function \( f(x) = \sum_{i=1}^{N} a_i y_i K(x, x_i) + b \) can be softened to a hard classification by a logistic transformation of the plane, and the probability of each cell being converted to a fire point can then be calculated as Equation (12).

\[
P_c = \frac{1}{1 + e^{-\left(\sum_{i=1}^{N} a_i y_i K(x, x_i) + b\right)}}
\]  

(12)

where \( x_i \) is the support vector, \( y_i \) is the y value corresponding to \( x_i \), \( P_c \) is the probability that the fire influence factor affects the cell transition, \( a_i \) is the Lagrangian coefficient of the \( x_i \) object, and \( K(x, x_i) \) is the kernel function. Here, a kernel function with a strong localization, such as the Gaussian radial basis kernel function is chosen, i.e., \( K(x, x_i) = e^{-\frac{1}{2\sigma^2}||x-x_i||^2} \).

Then Equation (12) can be rewritten as Equation (13)

\[
P_c = \frac{1}{1 + e^{-\left(\sum_{i=1}^{N} a_i y_i e^{-\frac{1}{2\sigma^2}||x-x_i||^2} + b\right)}}
\]  

(13)

where \( x_k \) is the vector corresponding to the kth cell, \( a_i \) is the Lagrangian coefficient of the \( x_i \) object, \( x_j \) is the support vector, \( y_i \) is the y-value corresponding to \( x_j \), the decision function penalty parameter is \( c \), and the radial basis function coefficients \( \sigma \) are selected by a combination of comparisons, with \( c = 1 \) and \( \sigma = 1 \). Then, Equation (13) can be rewritten as Equation (14).

\[
P_c = \frac{1}{1 + e^{-\left(\sum_{i=1}^{N} a_i y_i e^{-\frac{1}{2\sigma^2}||x-x_i||^2} + b\right)}}
\]  

(14)

However, the use of meta-cellular automata for modeling forest fires cannot be simply regarded as a classification problem. The various interactions within the meta-cellular neighborhood and the influence of adjacent wind effects must also be considered.

According to the work of GA Trunfio [59,63], cell transition functions can be affected by two parts: internal transfer and local interactions. For each cell \( A(i,j) \) in this study, the cell state transfer probability can also be expressed as a function of two subcomponents, as shown in Equation (15).

\[
P_{t,(i,j)} = p(i,j)_t \times \theta(i,j)_t
\]  

(15)

where \( p(i,j)_t \) is the ignition probability (Equation (13)), which is used to measure the ignition probability of the cell affected by its own factors, and \( \theta(i,j)_t \) represents the adjacent wind effect on this cell. Hence, the formula can be changed to Equation (16):

\[
p_{t,(i,j)} = \frac{1}{1 + e^{-\left(\sum_{i=1}^{N} a_i y_i e^{-\frac{1}{2\sigma^2}||x-x_i||^2} + b\right)}} \times \frac{\sum_{i=0}^{n} \sum_{j=0}^{n} K_{w}(i,j)}{\sum_{i=0}^{n} \sum_{j=0}^{n} K_{w}}
\]  

(16)

2.4. Spreading Time Step

Another key element for forest fires is the time step, which influences the spatial distribution characteristics of forest fire spread. The time step \( t \) is calculated using the following Equation (17).

\[
t = \frac{L}{R}
\]  

(17)

where \( L \) is the centroid distance between two adjacent cells. \( R \) is the rate of fire spread considering the spatial fuel distribution and influence of wind and terrain slope, which is evaluated multiplying the initial fire spread speed by the fuel spatial distribution and arrangement adjusted index, by the wind speed adjusted index \( (K_w) \), and by the slope adjusted index \( (K_{\phi}) \) [74,78,79].
where $\phi$ is the terrain slope angle.

With this, the rise of the fire front isometrics can be simulated with the iteration at each time step throughout the burning process. The continuous time of the fire spreading process can be converted into a discrete-time state using the time step such that the states of fire can be updated during the fire propagate process at each time step.

3. Case Studies

3.1. Study Area and Data

The historical fire data of the Lushan area of Liangshan Prefecture, Sichuan Province, were collected for use in verifying the proposed model (Figure 4). A point source ignition is used in this scenario—which was investigated and recognized as the actual forest fire starting point—that broke out in this area at 15:35 on 30 March 2020 just before the major fire spread event started, causing a total area of 3047.7805 hectares of various types of burned land, and 791.6 hectares of fire-affected forest area [80]. Lushan Mountain is 2317 m above sea-level, bordered by Qionghai Sea in the east, Anning River to the west, the ancient city of Xichang to the north, and the Luoji Mountain to the south. The wind power on the day of the fire reached 7–8 [80]. Using ArcGIS tools, the influencing factors, slope, direction, elevation, NDVI, and relative humidity were obtained from the remote sensing data for the area.

![Figure 4. Study area.](image)

First, the dependent variable data (fire point) and independent variable data (slope, aspect, elevation, relative humidity, normalized vegetation index) of Xichang and Lushan areas were obtained, and then 30% of Xichang data were screened through SPSS and 70% of the Lushan data were used as training data for the least squares support vector machine. Then, the least squares support vector machine was used for classification processing, and after training using 25,133 sample points, 25,133 support vectors were obtained, and with $b = -0.9512$ and Equation (14), the final fire impact probability expression can be determined as Equation (19).

$$P_c = \frac{1}{1 + e^{-\left(\sum_{i=1}^{N} a_i y_i e^{\frac{1}{2}||x_i - x_j||^2} - 0.9512}\right)}}$$  

The final fire ignition probability can be obtained by using Equation (13) and the adjacent wind effect. Then, the cellular automata can be used to simulate the spreading of fire.

3.2. Simulation Results and Analysis

Figure 5 shows the simulation of the fire spreading process from the initial fire point to the end every 5 h. For each model, the simulated fire spread direction is basically in
line with the actual fire spread direction of official reports [80]. From Figure 5, it can be found that the fire spread rate in the northeast direction is faster than that in the southwest direction due to the influence of wind. However, there is dense vegetation in the southwest direction, where there are burning conditions. Although the southwest direction is not downwind, there is also a tendency for fire spreading in this direction. The proposed method and the basic CA simulation of fire spread and the actual spread direction are shown in Figure 5b,c, respectively. Finally, the simulated burned area of the forest fire can be obtained. Comparing the simulated burned area with the actual burned area (Figure 6), the difference between the simulated forest fire area and the actual area can be seen more clearly, as shown in Figure 7.

![Simulation of fire spread](image)

**Figure 5.** Simulation of the entire forest fire process in 5 h intervals.
Figure 6. The actual burned area of the forest fire.

Figure 7. Comparison between the actual fire area and the simulation results.
In order to more accurately simulate and observe the trend of forest fire spread, it is necessary to evaluate the spatial accuracy of the results of the spread model. In forest fires, the burned area is important for evaluating the fire size, estimating the loss of stockpile, and calculating the economic loss due to the fire. Equations (20)–(22) are commonly used to evaluate the space accuracy of the fire spreading model [81].

\[
\mu = \frac{S_1 \cap S_2}{S_2} \times 100\% \tag{20}
\]

\[
\varepsilon_1 = \frac{S_2 - (S_1 \cap S_2)}{S_2} \times 100\% \tag{21}
\]

where \(S_1\) is the simulated burned area, and \(S_2\) is the actual fire area. \(\mu\) is the ratio of the area where there is overlap between the simulated burned area and the actual fire area to the actual fire area. \(\varepsilon_1\) is the ratio of the actual fire area that is not predicted to the actual fire area. \(\mu\) and \(\varepsilon_1\) describe the degree of intersection between the simulated and the actual fire area.

\[
\varepsilon_2 = \frac{S_1 - (S_1 \cap S_2)}{S_2} \times 100\% \tag{22}
\]

\(\varepsilon_2\) is the ratio of the simulated burned area that was not actually burned to the actual fire area. It describes the deviation between the simulated burned area and the actual fire in spatial terms.

The actual fire area is 33,735 grids (Figure 6). In terms of the burned area in Figure 5, as shown in Table 1, the number of burned grids simulated by LSSVM is 36,111 grids. The number of simulated burned grids is 33,793 and 33,725 grids in ELM-CA and the basic CA model, respectively. From Table 1, we can see that the simulated burned area determined using the three models mostly coincides with the actual fire area. Compared with the other methods, the proposed method demonstrates lower deviation between the simulated burned area and the actual fire in spatial terms.

<table>
<thead>
<tr>
<th>State</th>
<th>LSSVM-CA</th>
<th>ELM-CA</th>
<th>CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu)</td>
<td>33,028 (97.9%)</td>
<td>30,602 (90.71%)</td>
<td>29,634 (87.84%)</td>
</tr>
<tr>
<td>(\varepsilon_1)</td>
<td>707 (2.1%)</td>
<td>3133 (9.29%)</td>
<td>4101 (12.16%)</td>
</tr>
<tr>
<td>(\varepsilon_2)</td>
<td>3083 (9.14%)</td>
<td>3191 (9.46%)</td>
<td>4091 (12.13%)</td>
</tr>
<tr>
<td>(S_2)</td>
<td></td>
<td></td>
<td>33,735</td>
</tr>
<tr>
<td>(S_1)</td>
<td>36,111</td>
<td>33,793</td>
<td>33,725</td>
</tr>
</tbody>
</table>

In order to further evaluate the performance of the model, two other forest fires that occurred in Liangshan Prefecture, Sichuan Province, were simulated. In recent years, fires have frequently occurred in Liangshan Prefecture. The Muli fire was the first fire subjected to simulation. It occurred on 28 March 2020 at the junction of Qiaowa Town and Xiangjiiao Village in Muli County. The fire was caused by human carelessness. The wind on the day of the fire was in the northwest direction and reached 2.7 m/s.

The second fire subjected to simulation was the fire near Baidiao country, which occurred on 3 April 2020. Most of the area is high mountains that are low in the south and high in the north, near the Yalong river. The wind on the day of the fire was in a southwest direction and was about 1.5 m/s.

Figure 8 shows the simulation of the initial spreading process of the Muli and Baidiao fires. For each fire, the simulated fire growth is basically in line with the spread direction of the actual fires. From Table 2, we can see burned areas simulated by the model basically coincide with the majority of the actual fire area (Figure 9). The results of the proposed method show good spatial accuracy between the simulated burned area and the actual fire.
Figure 8. The simulated processes of the Muli and Baidiao fires.

Figure 9. Comparison between the actual fire area and the simulation results in Muli and Baidiao fires.
Table 2. The results of simulated Muli and Baidiao fires.

<table>
<thead>
<tr>
<th>State</th>
<th>Muli</th>
<th>Baidiao</th>
</tr>
</thead>
<tbody>
<tr>
<td>µ</td>
<td>4091 (83.85%)</td>
<td>5408 (85.0%)</td>
</tr>
<tr>
<td>ε₁</td>
<td>788 (16.15%)</td>
<td>957 (15.0%)</td>
</tr>
<tr>
<td>ε₂</td>
<td>787 (16.13%)</td>
<td>350 (5.5%)</td>
</tr>
<tr>
<td>S₁</td>
<td>4878</td>
<td>5758</td>
</tr>
<tr>
<td>S₂</td>
<td>4879</td>
<td>6365</td>
</tr>
</tbody>
</table>

4. Discussion

The spread of forest fire contains inherent uncertainties due to the diversity in combustible material and the complexity of the geographic environment. It is not easy to obtain the exact parameters of various combustible material characteristics of the whole forest area or to apply localized fuel behaviors to the whole area [64,69].

Machine learning-based CA models are attractive for application in simulating forest fire spread as they allow spatiotemporal modeling of the complicated spread mechanisms without the need for an in-depth preliminary study of their general evolutionary causes [63,82]. With the increase in computerized modeling, CA have been commonly used to simulate fire spread as discrete processes on a regular spaced cell [62,68,69]. Although good simulation performance was obtained in the LSSVM-CA model, it is not without limitations.

In a fire, the combustion material in the burning cell may affect the surrounding non-adjacent cells in an unintended way [83]. For example, the burning material may fly into and ignite the non-adjacent cells due to blowing wind. This phenomenon could cause underestimation of the real fire spreading speed in proposed models [82,84,85].

The state of a cell is affected by transition rules and the state of its adjacent cells [86]. When a central cell has multiple adjacent burning cells to which the fire spreads at the same time, the contribution to its state transition may be underestimated. Consequently, this will impact the model effectiveness [85].

It is not easy to compare the performance of one model with another when different fire characteristics and fuel are employed or different kinds of transition rules (such as stochastic and deterministic) are used. However, some general observations of CA-based methods can be obtained on the basis of the transition rules [82].

In [85], the state weight coefficient of a burned cell was introduced to account for the overlapping of the burning areas, then an empirical fire spread model was used. However, the use of a wind coefficient resulted in simplification of the effect of the adjacent wind.

Due to the complexity of the forest environment, the factors and parameters involved in modeling of a specific forest environment may not be available in other places. Thus, a model designed for a specific local area may not be universal and, thus, cannot be directly applied elsewhere [86]. In this model, we adopted widely used factors, such as terrain, climate, and vegetation for modeling, measure of which are obtainable through publicly available data. This makes the modeling method more applicable.

Forest fire is a multistage phenomenon taking into account inert heating, drying, pyrolysis, ignition, flame combustion, and coke after burning. The behavior at the fire front is mainly dependent on the fire properties of the fuel; however, some dynamic inputs and factors may not be directly obtained in CA [87,88]. For instance, in order to select the available model, up to 15 kinds of the rate of fire spread models are trained to select the most appropriate model based on a burning experiment [89].

Considering the uncertainties, empirical/statistical and physical/theoretical models are helpful for analyzing and simulating fire behavior [90]. Thus, CA modeling is required to adopt some equations of the fire behavior models to adapt this dynamic process [87,91,92]. To do so, statistical models and theoretical models are commonly integrated into the CA framework. However, some models require numerous primary inputs involving many factors. This may reduce their applicability.
5. Conclusions

In this paper, a new fire spread modeling method, LSSVM-CA, is proposed, in which the basic framework of traditional forest fire CA is combined with LSSVM classification. Moreover, the proposed method was validated based on historical fire data in the Lushan area of Xichang City, Liangshan Prefecture, Sichuan Province. The results show that using the least squares support vector machine model, the forest fire ignition probability formula can be effectively derived. The proposed modeling method demonstrated good performance in accurately simulating fire spread. Although the proposed LSSVM-CA model achieved relatively good performance, there are some issues to be considered in future work.

Vegetation is widely regarded as the fuel of a fire. Among all the factors (such as terrain, climate, etc.), that are used for fire spread simulation, vegetation information thus plays an important role in fire spread. With the developments in measurement technology, airborne light detection and ranging data could be used to derive and map forest attributes, such as vegetation type, density, and volume. Such forest attributes could be used as input factors to improve model performance.

Due to the historically frequent occurrence of fire in this area, it is urgent to design effective fire management strategies. In the future, we will use this model to simulate and obtain potential key fire propagation paths so as to plan efficient forest fire extinguishing strategies. These strategies include optimizing the allocation of fire-fighting resources and expanding road networks, such that the fire extinguishing capacity of the whole area can be improved at minimum cost.

Due to the complexity of fire spread, adopting and combining the most appropriate mathematical fire behavior model for a specific area into the CA modeling may improve the performance to some extent. In the future, we would investigate and try to combine CA with equations that can adequately reflect the relationships between fire behavior and the numerical values of established conditions, such as ignition risk, thermal conductivity, and flame height.

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

Funding: This work was supported in part by the Open Foundation of Industrial Software Engineering Technology Research and Development Center of Jiangsu Education Department [201050621ZK004] and the Start-up Fund for New Talented Researchers of Nanjing Vocational University of Industry Technology [Grant No. YK22-05-03].

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References


10. Matin, M.A.; Chitale, V.S.; Murthy, M.S.; Uddin, K.; Bajracharya, B.; Pradhan, S. Understanding forest fire patterns and risk in Nepal using remote sensing, geographic information system and historical fire data. Int. J. Wildland Fire 2017, 26, 276–286. [CrossRef]


27. Cruz, M.G.; Alexander, M.E. Uncertainty associated with model predictions of surface and crown fire rates of spread. Environ. Model. Software. 2013, 47, 16–28. [CrossRef]


33. Rothermel, R. How to Predict the Spread and Intensity of Forest Fire and Range Fires; General Technical Reports, INT-143; US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station: Ogden, UT, USA, 1983.

