Received 13 March 2024, accepted 11 April 2024, date of publication 16 April 2024, date of current version 24 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3389035

RESEARCH ARTICLE

A Forest Fire Prediction Model Based on Cellular Automata and Machine Learning

XUAN SUN[1], NING LI[2], DUOQI CHEN[3], GUANG CHEN[1], CHANGJUN SUN[1], MULIN SHI[4], XUEHONG GAO[5], KUO WANG[5], AND IBRAHIM M. HEZAM[6]

[1]China Academy of Building Research, Beijing 100029, China
[2]Xinjiang Institute of Technology, Xinjiang 830099, China
[3]Fire Rescue Detachment, Tongzhou, Beijing 101100, China
[4]Binzhou Safety Science and Technology Service Center, Binzhou 256600, China
[5]School of Civil and Resource Engineering, University of Science and Technology Beijing, Beijing 100083, China
[6]College of Sciences, King Saud University, Riyadh 11362, Saudi Arabia

Corresponding author: Ning Li (lining2023@xjit.edu.cn)

This work was supported by the Fundamental Research Funds for the Central Universities under Grant FRF-IDRY-21-016 and King Saud University through Researchers Supporting Project number (RSP2024R389).

ABSTRACT Forest fires constitute a widespread and impactful natural disaster, annually ravaging millions of hectares of forests and posing a severe threat to human life and property. Accurate quantitative prediction of forest fire spread is essential for devising swift risk management strategies and implementing effective firefighting approaches. In response to this imperative, this study introduces a Forest Fire Spread Behavior Prediction (FFSBP) model, encompassing two integral components: the Forest Fire Spread Process Prediction (FFSPP) model and the Forest Fire Spread Results Prediction (FFSRP) model. The FFSPP model involves the prediction of the direction and speed of forest fire spread, achieved through a fusion of the Cellular Automata model and the Wang Zhengfei model. On the other hand, the FFSRP model focuses on forecasting the extent of the burned area, leveraging machine learning methods. To validate the efficacy of the proposed models, a real case study is undertaken using the "3.29 Forest Fire" incident in China. The FFSPP model is validated against this case, while the FFSRP model is assessed using a real fire dataset obtained from Montesinho National Forest Park in Portugal. Results from the validation process reveal that during the natural development period of the "3.29 Forest Fire," the FFSPP model predicts a burned area of 286.81 hm^2, with an associated relative error of 28.94%. This relative error is notably smaller than those observed in the Farsite and Prometheus fire behavior simulation models. Additionally, the FFSRP model demonstrates commendable predictive performance, particularly in the context of small and medium-sized fire scenarios. These findings underscore the potential of the FFSBP model as a valuable tool in enhancing forest fire prediction accuracy and facilitating more robust risk mitigation and firefighting strategies.

INDEX TERMS Forest fire prediction, forest fire spread, cellular automata, Wang Zhengfei model, machine learning.

I. INTRODUCTION

The escalating frequency and magnitude of global forest fires, exacerbated by factors like global warming and heightened extreme weather events, pose severe threats to human lives, property, and ecological environments [1], [2]. Annually, millions of hectares of forests succumb to flames, incurring substantial financial costs and casualties. Accurate prediction of forest fires plays a pivotal role in averting numerous calamities. Forest fire prediction typically falls into three categories: fire risk weather prediction, forest fire occurrence prediction, and forest fire behavior prediction. These categories consider specific factors: (i) fire risk weather prediction focuses on meteorological elements [3], [4]; (ii) forest fire occurrence prediction involves meteorological factors, combustibles, and fire sources [5], [6], [7]; (iii) forest fire behavior prediction incorporates meteorological factors, combustibles, and terrain conditions [8], [9], [10]. While weather and occurrence predictions assess the potential for, or likelihood of, forest fires, behavior prediction, encompassing weather, combustibles, and terrain, delves into the direction, speed of spread, and burned area—crucial for

The associate editor coordinating the review of this manuscript and approving it for publication was Wojciech Sałabun.

effective firefighting [11], [12]. This study concentrates on forest fire behavior prediction, forecasting both the process and outcomes of forest fire occurrences. The predicted data for the occurrence process includes direction and speed of fire spread, visually presented. Outcome predictions encompass the anticipated burned area.

Forest fire spread, a facet of forest fire behavior, pertains to the characteristics exhibited by combustibles from ignition to extinction. Employing mathematical methods under simplified conditions, the forest fire spread model establishes quantitative relationships between key parameters (e.g., fuel properties, terrain, meteorological factors) and forest fire behavior, including spread speed [13]. These relationships facilitate the prediction of impending or ongoing forest fire behavior, guiding firefighting and daily forest management. Since W.R. Fons introduced a mathematical model in 1946, scholars worldwide have proposed various models based on different assumptions for combustible materials. Notable models include the Canadian forest fire spread model [14], the Australian McArthur model [15], the American Rothermel model [16], [17], the Chinese Wang Zhengfei forest fire spread model [18], and the modified models based on these models [18], [19], [20]. Despite their utility, each model has limitations, especially when assumptions are absent, leading to potential errors. Therefore, understanding the applicability, conditions, and pros and cons of a model is crucial before utilization. These models fall into empirical, physical, and semi-empirical/semi-physical categories based on principles. Empirical models rely on statistical analysis of actual data without considering physical mechanisms. Physical models are based on energy conservation and heat conduction laws but may face challenges in collecting input parameters. Semi-empirical/semi-physical models, like the Wang Zhengfei and Rothermel models, incorporate experimental data guided by specific physical mechanisms. However, the 11 input items of Rothermel model with complex relationships and practical acquisition requirements pose challenges to its application.

Traditional forest fire spread models, often grounded in mathematical or physical laws, lack inherent self-organizing mechanisms. As system complexity rises or disturbances occur, the challenge of solving differential equations or estimating numerical values intensifies [21], [22]. The introduction of Cellular Automata (CA) addresses this limitation, offering a remedy for the absence of self-organization in conventional models and providing a more vivid depiction of forest fire spread [13], [23], [24]. To create a user-friendly and self-organizing model, this study integrates the Wang Zhengfei model with the Cellular Automata (CA) model. This integrated approach aims to predict the direction and speed of forest fire spread, ultimately achieving a visually intuitive representation.

Furthermore, as computer computing power has advanced and Machine Learning (ML) has evolved, ML has garnered attention for its ability to discern nonlinear relationships among diverse input parameters. Researchers have applied ML to forest fire prediction, utilizing methods such as backpropagation neural networks, random forests (RFs), deep learning, and ensemble learning [6], [24], [25], [26]. While various ML techniques have demonstrated superior results in fire risk prediction compared to probability and statistical methods, the emphasis has primarily been on predicting the likelihood of forest fire occurrence rather than the burned area. In recent decades, ensemble learning has gained prominence in the ML field due to its efficacy in addressing practical application challenges [26], [27], [28]. Consequently, this article opts for ensemble learning to forecast the outcome of forest fire occurrences, specifically, the burned area. This choice aims to enhance the prediction of the fire situation and facilitate timely response measures.

In this context, the focus of this paper is on investigating the prediction of both the forest fire spread process and its outcomes, leading to the establishment of a Forest Fire Spread Behavior Prediction (FFSBP) model. This model comprises two integral components: the Forest Fire Spread Process Prediction (FFSPP) model and the Forest Fire Spread Results Prediction (FFSRP) model. The FFSPP model involves combining the Wang Zhengfei model with the Cellular Automata (CA) model to predict the direction and speed of forest fire spread, while the FFSRP model employs ensemble learning methods to predict the burned area. To validate the forest fire spread prediction model, the "3.29" forest fire in Anning, Southwest China, serves as an illustrative example. Additionally, the burned area prediction model is verified using a real fire dataset from January 2000 to December 2003 obtained from Montesinho National Forest Park in Portugal. The research outcomes hold significant practical implications for forest fire response: (i) predicting and visualizing the direction and speed of forest fire spread offers valuable insights for deploying firefighting resources; (ii) forecasting the burned area provides guidance for aggregating total firefighting resources; (iii) the proposed method's straightforward calculation process, coupled with its adaptability to different datasets, facilitates easy application and optimization of the model.

The rest of this paper is organized as follows. Section II critically reviews previous studies pertaining to the primary concerns of this research, namely forest fire impact factors, forest fire spread prediction, and burned area prediction, with a focus on identifying research gaps. In Section III, the FFSBP model is introduced, along with the development of the corresponding methodology. Additionally, Section IV presents a case study featuring computational results to validate the effectiveness of the proposed model. Finally, Section V concludes the research by offering valuable insights into forest fire prediction and delineating potential future directions.

II. LITRTATURE REVIEW

A. FOREST FIRE IMPACT FACTORS

Numerous studies have investigated the influencing factors of forest fires. Wu, Li [29] conducted a study on the driving factors of forest fires in different provinces of China. They comprehensively considered anthropogenic factors, meteorological conditions, topography, and vegetation. Additionally, they established an artificial neural network model to predict the probability of forest fire occurrences in the research area. Guo et al. [30] applied the classic logistic regression model and the geographically weighted logistic regression model to ascertain the relationship between human-induced forest fires in northern China and their potential driving factors. The research findings confirmed the significance of distance from railways, elevation, fire line length, and vegetation coverage in the occurrence of forest fires in northern China. Preisler and Westerling [31] studied the strongly impact of meteorological factors such as air temperature, precipitation, and humidity on the occurrences and dynamics of fire. College of Forestry, Forestry University [32] conducted research on the trends and driving factors of forest fires based on MODIS satellite fire point data, integrating meteorological, human, topographical, and vegetation factors. Trend analysis and a Logistic regression model were employed in the study. Ma et al. [33] analyzed the impacts of climate, topographic, vegetation and socioeconomic variables on forest fire occurrence in six geographical regions in China. The results show clear regional differences in the forest fire driving factors and their impacts in China. Among them, climate variables are the forest fire driving factors in all regions. Li et al. [34] investigated the quantitative effects of factors such as forest location, species type, fire occurrence date, temperature, and wind speed on the degree of forest fire disaster. The results indicate that forest location has significant impacts on forested area burned. In summary, the principal factors influencing forest fires encompass the combustible state, meteorological factors, and topographical conditions. However, existing research on predicting forest fire behavior often falls short in either integrating all these factors or relies heavily on long-term historical data, resulting in a lack of comprehensive and specific predictions. Consequently, this paper aims to bridge this gap by integrating the aforementioned three factors and utilizing short-term historical data for a more precise prediction of forest fire behavior.

B. FOREST FIRE SPREAD PREDICTION MODEL BASED ON CA

It is well known that the CA is characterized as a dynamic system undergoing evolution within a discrete, finite state cellular space, guided by specific rules in a discrete time dimension, which has undergone thorough exploration and widespread application. Besides, the CA modeling method is straightforward and practical, enabling the prediction of forest fire behavior from a microscopic perspective. The algorithm's swiftness adds to its appeal. Aleixo et al. [35] employed site percolation and SIR epidemiology rules within a CA to simulate local fire dynamics. Notably, phase transitions were identified for various combinations of fire risk within each class, and these values were utilized to parameterize the resulting landscape network. Hui, Rui [36] proposed a simulation algorithm that couples a geographic CA to address the issues of high errors and low efficiency in traditional forest fire spread simulation models when simulating large-scale forest fires. Mahdizadeh and Navid [37] utilized a CA model to simulate the spread of wildfires. The model considered the most influential spatial and temporal driving factors for wildfire propagation, including wind speed and direction, vegetation type and density, as well as topographical conditions. Jellouli et al. [38] explored the application of CA methods for simulating forest fire phenomena. The model considered key parameters such as natural vegetation, density, humidity, wind force, and elevation. Li et al. [39] proposed a forest fire spread simulation model using CA with long short-term memory (LSTM) based on the interaction between wind and fire. Xu et al. [40] propose a new method combining least squares support vector machines (LSSVM) with forest fire CA framework, namely LSSVM-CA, in which the effects of adjacent wind on the law of fire spread are considered. Although these models perform well in predicting the spread of forest fires, the single forest fire factors considered may not be effective in practical applications. Besides, the factors considered in a given study are contingent upon the data accessibility for the specific application region. Given the inherent complexity arising from the multitude of factors, applying the model to real-case scenarios may pose challenges. Consequently, this paper aims to enhance predictive accuracy by integrating the Wang Zhengfei model with CA. It selects easily obtainable factors, including combustible state, meteorological conditions, and topography, to predict the direction and speed of forest fire spread and visualize the results.

C. FOREST FIRE BURNED AREA PREDICTION BASED ON ML

In recent years, ML has gained widespread attention for its capacity to discern nonlinear relationships among diverse parameters. Consequently, ML methods have been increasingly applied to predict the burned area of forest fires. Frédéric et al. [41] proposed a hybrid architecture deep neural network that can simultaneously process different types of input data for estimating fire spread under different environmental conditions. The input data of the model includes two-dimensional images of the surrounding landscape and combustion parameters, and the final output is the fire area. Cortez and Morais [42] proposed a support vector machine regression model to predict the area of forest fires. The input data of this model includes temperature, humidity, wind speed, and precipitation. The radial basis kernel function and ϵ-insensitive loss function are used to optimize and obtain a complete model for predicting the area of fire. Bisquert et al. [43] selected 12 forest fire impact factors to form a feature vector, established a 3-layer backpropagation neural network to predict the burned area of forest fire, and determined the fire risk level based on the burned area of fire.

Nevertheless, existing studies exhibit limitations in accurately predicting the burned area of forest fires. Certain models, as illustrated by Frédéric et al. [41], incorporate

TABLE 1. Summary of the literature pertaining to pedestrian evacuation.

Article	Main problem	3-type influential factors	Parameters easily obtained or not	Real time/short-term forecasting	Model	Approach
Nikolaos, Marios [3]	Fire risk weather prediction	No	Yes	No	Weather prediction model	ML; Canadian fire weather index
Deng, Zhang [4]	Fire risk weather prediction	No	Yes	No	RegCM model	RCM; WRF
Pang, Li [5]	Forest fire occurrence prediction	Yes	No	No	Occurrence prediction model	RF; ML
Chao, Honglei [6]	Forest fire occurrence prediction	Yes	No	No	Occurrence prediction model	RF BPNN; LR
Shi and Zhang [7]	Forest fire occurrence prediction	Yes	No	No	PSO-RF model	RF; PSO; SVM; LR
N.S., H.P. [10]	Forest fire behavior prediction	Yes	No	Yes	AdaBoost model	RF; ANN; SVM;
Denham, Wendt [11]	Forest fire behavior prediction	No	No	Yes	DDDGA model	Genetic algorithm
Ntinas, Moutafis [12]	Forest fire behavior prediction	No	No	Yes	FCA model	CA; Parallelization
Xu, Li [40]	Forest fire behavior prediction	No	No	Yes	LSSVM-CA model	LSSVM; CA
Aleixo, Clara [35]	Forest fire behavior prediction	No	No	Yes	Percolation model	CA; Multi-scale network
Li, Zhang [39]	Forest fire behavior prediction	No	No	Yes	LSTM-CA model	LSTM; CA; Extreme learning machine
Frédéric, Vivien [41]	Forest fire behavior prediction	Yes	No	Yes	Hybrid architecture	Deep neural network; Numerical simulation
Bisquert, Caselles [43]	Forest fire occurrence prediction	No	Yes	No	Occurrence prediction model	Artificial neural networks
This study	**Forest fire spread behavior prediction**	**Yes**	**Yes**	**Yes**	**FFSBP model**	**CA; Wang Zhengfei model; ML**

intricate input data, such as combustion parameters, which are often challenging to obtain. Conversely, other models [42], [43] grapple with an excess or insufficiency of input data, leading to either an increased data collection workload or suboptimal prediction performance during application. Recognizing the potential of ensemble learning, which enhances model performance by amalgamating multiple ML models, especially when confronted with intricate data and tasks, this paper adopts ensemble learning methods. The approach involves combining multiple learners to predict the burned area of forest fires, incorporating multicollinearity tests for data screening and grid search methods for model optimization.

D. SUMMARY

In addition, a comprehensive summary of studies in recent years is provided in Table 1, where the main concerns of the FFSBP model are compared with the previous studies to illustrate the research gaps and novelties of this study. As presented in Table 1, what emerges from the recent literature survey is that CA and ML methods can effectively predict forest fire behavior. Still, many studies only consider a portion of the factors affecting forest fires. Besides, the input data for most forest fire spread models is hard to obtain in practical applications. Finally, the research on the burned area of forest fires mainly focused on monitoring which only provides fire information from a certain moment in the past.

According to the main concerns and research gaps identified in Table 1, this study makes three main research contributions. First of all, this study uses CA and ML methods to predict forest fire behavior, aiming to enhance firefighting efforts for the preservation of life and property. Notably, the factors influencing forest fires, particularly combustible state, meteorological conditions, and topography, are integral components of this predictive approach. Besides, this study integrates the CA model with the Wang Zhengfei model to predict the direction and speed of forest fire spread, providing a visual representation of the results. Furthermore, the ML method was used to predict the burned area of forest fires, including using multicollinearity tests to screen data and grid search methods to optimize the model. Subsequently, a real-case study is conducted using the "3.29 Forest Fire" in China and a genuine fire dataset from Montesinho National Forest Park in Portugal, to validate the efficacy of the proposed methodology.

III. METHODOLOGY

A. FFSPP MODEL

1) MAIN SETTINGS OF CA

The forest fire spread model based on CA is a discrete event model that divides forest land into numerous discrete cells, simulating the fire spread process according to a defined set of rules [25], [44]. Each cell denotes a location on the forest land, and the model simulates the fire propagation by

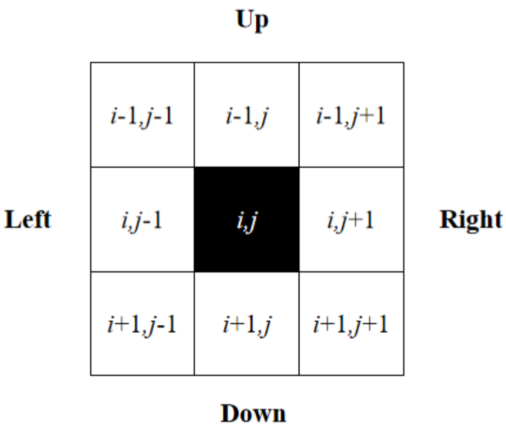

FIGURE 1. Moore-type neighboring cells.

updating the status of each cell. The key settings for a forest fire spread model based on CA include:

(1) State definition: Each cell can possess distinct states. This paper adopts four states: "unburned," "partially burned," "fully burned," and "end burned." Initially, the fire source status is assigned to one or more cells based on the actual situation.

(2) Proximity definition: The adjacent cells of each cell are defined. Typically, four or eight neighborhoods represent cells adjacent to each other, with this paper opting for eight neighborhoods to accurately represent units adjacent at the top, bottom, left, right, or diagonal.

(3) Update rules: Rules for updating the status of each cell are defined based on the status of adjacent cells and environmental factors like wind direction, wind speed, fuel humidity, and terrain. For instance, when a cell is in a burning state, surrounding unburned cells can be ignited.

(4) Iterative update: The spread of fire is simulated by iteratively updating the status of each cell. Discrete time steps control the frequency of updates. Each update considers the current state and the state of surrounding cells, updating according to predefined rules.

(5) Boundary condition handling: Special handling of boundary conditions is required for boundary cells to prevent the fire from spreading beyond the simulation range.

2) IMPROVEMENT OF CA FOR FOREST FIRE SPREAD

The fundamental components of CA encompass four elements: cell, state, neighborhood, and rule. Presently, research predominantly focuses on one-dimensional and two-dimensional CA. This study opts for a two-dimensional CA, akin to a regular grid. Each grid serves as a cell, possessing its state at every moment. The transition of cell states is contingent upon dynamic rules: functions reliant on the current state of the cell and its neighboring cells to ascertain the subsequent state. Throughout the simulation process, dynamic iterations and calculations are executed, incorporating changes in the neighborhood based on the transformation rules.

TABLE 2. Partial notations used in this study.

notations	significance	unit
R	the speed of forest fire spread	m/min
R_0	initial spreading speed in the absence of wind	m/min
K_s	combustible correction factor	
K_w	wind speed correction coefficient	
$K_\emptyset$	terrain slope correction factor	
$\emptyset$	terrain slope angle	
T	daily maximum temperature	°C
W	the average wind speed level at noon	grade
h	the daily minimum relative humidity	%
V	wind speed	m/s

TABLE 3. Wind speed level W table.

W	Wind speed (m/s)	W	Wind speed (m/s)	W	Wind speed (m/s)
0	0.0-0.2	6	10.8-13.8	12	32.7-36.9
1	0.3-1.5	7	13.9-17.1	13	37.0-41.4
2	1.6-3.3	8	17.2-20.7	14	41.5-46.1
3	3.4-5.4	9	20.8-24.4	15	46.2-50.9
4	5.5-7.9	10	24.5-28.4	16	51.0-56.0
5	8.0-10.7	11	28.5-32.6	17	≥56.1

TABLE 4. Combustible correction factors K_s.

Fuel type	K_s	Fuel type	K_s
Flat needle leaf	0.8	Cypress/cherry-birch	1.8
Litter layer	1.2	Pasture grassland	2
Weed	1.6	Pinus Koraiensis/armandi/yunnanensis	1

In two-dimensional CA, rules are defined within a local spatial range, meaning that the state of a cell at the next moment is determined by its own state and the state of its neighboring cells. Therefore, before specifying rules, it is necessary to define certain neighbors and clarify which cells belong to the neighbors of that cell. This paper adopts Moore type neighborhood.

In a Moore-type neighborhood, the four adjacent cells on top, bottom, left, and right of a cell, along with four sub adjacent cells in the diagonal direction, are the neighbors of that cell. As shown in Fig. 1, adjacent cells are cells that share a common edge with the central cell (i, j), represented by (i-1, j), (i+1, j), (i, j-1), and (i, j+1), respectively. The sub adjacent cells are four positions located diagonally, represented by (i-1, j-1), (i-1, j+1), (i+1, j-1), and (i+1, j+1), respectively. The state of the cell (i, j) at time t in this paper is defined as (1), as shown at the bottom of the next page:

Its value range is: $0 \leq A_{ij}^t \leq 1$. If $A_{ij}^t = 0$, indicating that the cell (i, j) is not burning at time t; $0<A_{ij}^t<1$ indicates partial

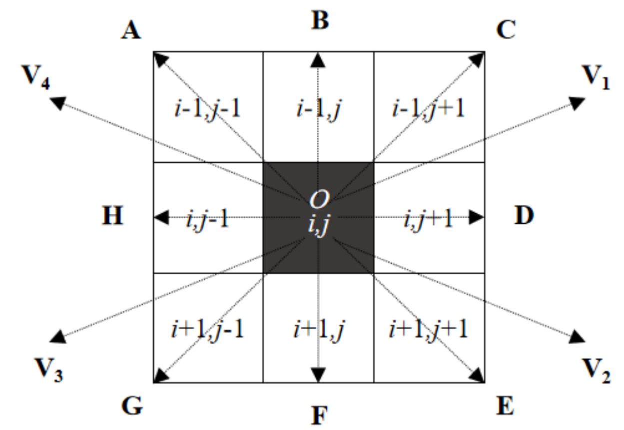

FIGURE 2. **Wind projection.**

combustion of the cell (i, j) at time t; $A_{ij}^t = 1$ indicates that at time t, the cells (i, j) are completely burned. This model assumes that only fully burned cells will propagate fire to neighboring cells.

On this basis, corresponding state expressions can be added according to the actual situation. When conducting simulation experiments, due to simulation needs, from the time t+1 after $A_{ij}^t = 1$, the state of the cell can be set to 2, indicating that the cell has completed combustion and is spreading towards surrounding cells from time t+1.

The primary factors influencing the propagation of forest fires encompass the combustible state, meteorological conditions, and topographical features. The conventional mathematical model delineating the spread rate of forest fires establishes a quantitative relationship derived from these intrinsic factors. In this study, rules for cell transformation are developed based on the Wang Zhengfei model [45], [46]. This involves eight categories of combustibles, meteorological factors comprising temperature, humidity, wind speed, and direction, and terrain factors primarily focusing on the slope of the forest area. Table 2 presents the notations utilized in this study for formulating the state of a cell.

The wind speed level W table is shown below:

The relevant formulas of the model are as follows:

$$R = R_0 K_s K_\varphi K_W \tag{2}$$

$$R_0 = 0.0299T + 0.047W + 0.009 \times (100 - h) - 0.304 \tag{3}$$

The values of K_s in the model are shown in Table 4:

In the CA model, the calculation functions of K_φ and K_W need to be further derived.

In cellular space, for any of the eight neighboring cells (k, l), there is a corresponding K_φ and their own slope values $\tan\varphi$ relative to the central burning cell. Therefore, the K_φ of neighboring cells (k, l) relative to the central combustion cells

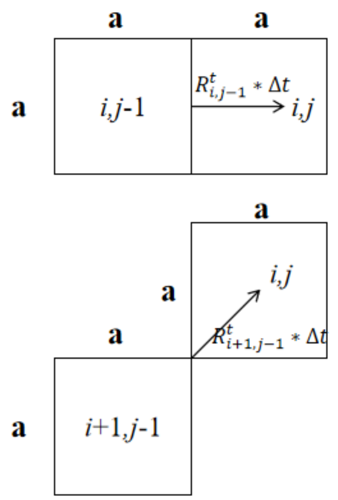

FIGURE 3. **The fire spread of the neighboring cell.**

(i, j) can be expressed as equations 4 or 5. If the cell (k, l) is an adjacent cell of cell (i, j), use equation 4; If the cell (k, l) is a sub adjacent cell of the cell (i, j), then use equation 5.

$$K_\varphi = e^{3.533(\tan\varphi)} = e^{3.5339(-1)^G\left|\frac{h}{a}\right|^{1.2}} \tag{4}$$

$$K_\varphi = e^{3.533(\tan\varphi)} = e^{3.5339(-1)^G\left|\frac{h}{\sqrt{2}\,a}\right|^{1.2}} \tag{5}$$

where h represents the height difference between the center positions of neighboring cells (k, l) and combustion cells (i, j), assuming that the height values within a cell are the same, equal to the height of the cell center point; a refers to the size of the cell's edge length; $\sqrt{2}a$ represents the diagonal length of the cell.

In the Wang Zhengfei model, $K_w = e^{0.1783V}$, which represents the relationship between K_W in the wind direction and wind speed V. There are eight neighboring cells in the cell space, so it is necessary to derive K_W value corresponding to the eight cells.

As shown in Fig. 2, $\overrightarrow{OA}$, $\rightrightarrows OB$, $\overrightarrow{OC}$, $\overrightarrow{OD}$, $\overrightarrow{OE}$, $\overrightarrow{OF}$, $\overrightarrow{OG}$, $\overrightarrow{OH}$ are the velocity directions of the central combustion cells (i, j) spreading towards the surrounding cells in eight directions; $\overrightarrow{OV_1}$, $\overrightarrow{OV_2}$, $\overrightarrow{OV_3}$, $\overrightarrow{OV_4}$ represent any direction in the four quadrants.

Then it can be seen that the propagation speed of the completely burned 8 central cells towards the surrounding, such as the propagation speed of the lower left corner cell ($\overrightarrow{\boldsymbol{OG}}$) can be expressed as: Then the propagation speed of the completely burned central cell towards the surrounding 8 cells can be obtained. For example, the propagation speed of the cell in the lower left corner ($\overrightarrow{\boldsymbol{OG}}$) is expressed by:

$$R_{i+1,j-1} = R_0 K_S K_W K_\varphi R_0 K_S e^{0.1783V\cos(225^\circ-\theta)}$$

$$A_{ij}^t = \frac{\text{Maximum combustion area affected by 8 directions of cells on cell } (i,j)}{\text{Area of cell } (i,j)} \tag{1}$$

$$\times e^{3.5339(-1)^G\left|\frac{h}{\sqrt{2}\,a}\right|^{1.2}} \tag{6}$$

As a bottom-up modeling method, CA can simulate the spatiotemporal dynamics of spatially complex systems by simply establishing basic transformation rules between neighboring local cells based on the ideal heat transfer law between burning and non-burning adjacent cells. Some researchers consider the change of a cell in cellular automaton from one state to another as the result of the influence of all factors at a certain moment [40]. In order to better fit the actual situation of fire spread, the transformation of cell state in this study is a continuous accumulation result - the result of all factors acting for a continuous period of time before a certain deadline. In addition, some researchers believe that adjacent cells and sub adjacent cells have the same impact on the central cell [24]. This study suggests that their impact is different and quantifies them differently. Determination of transformation rule function are as follows:

The combustion state of cell (i, j) at time t+1 is co-determination by the speed of its neighboring cell spreading to it and the combustion state of cell (i, j) at time t. The unburned or already burned cells at time t have no effect on the state of cells (i, j) at time t+1. Only fully burned cells can propagate in all directions at a speed. For example, if a fire spreads from the fully burned neighboring cell (i, j-1) on the left to the cell (i, j), and its propagation speed to the cell (i, j) is $R^t_{k,j}$, then within time t, due to the collinearity of neighboring cells (i, j-1), the combustion area of cell (i, j) is $aR^t_{i,j-1}\Delta t$, and the combustion area ratio is $\frac{aR^t_{i,j-1}\Delta t}{a^2} = \frac{R^t_{i,j-1}\Delta t}{a}$. Specifically, in the scenario where the fire spreads from the fully burned secondary cell (i+1, j-1) in the lower left corner to the cell (i, j), the combustion area ratio is $\frac{3.14\left(R^t_{i+1,j-1}\cdot\Delta t\right)^2}{4a^2} = \frac{0.785\left(R^t_{i+1,j-1}\cdot\Delta t\right)^2}{a^2}$. Similarly, if there are also completely burned cells in the other 6 cells, calculate their contribution to the combustion area of cells (i, j) using a similar method.

If $\boldsymbol{R^t_{I,J}}$ represents the speed at which fire spreads from cell (I, J) to cell (i, j) at time t, then the transformation rule function for updating the cell state can be expressed in the following after time Δt:

$$A^{t+1}_{i,j} = A^t_{i,j} + max\left\{\frac{max\left(R^t_{i-1,j}, R^t_{i,j-1}, R^t_{i+1,j}, R^t_{i,j+1}\right)\Delta t}{a}, \frac{0.785 * \left(max\left(R^t_{i-1,j-1}, R^t_{i-1,j+1}, R^t_{i+1,j+1}, R^t_{i+1,j-1}\right)\right)^2 \Delta t^2}{a^2}\right\} \tag{7}$$

If $A^{t+1}_{i,j} < 1$, then the fuel part in cell (i, j) burns, and the fire does not propagate to neighboring cells; If $A^{t+1}_{i,j} \geq 1$, the cell completely burns and begins to spread towards the surrounding 8 cell fires. A drawback of this approach is that the fire spread should not exceed the edge length of one cell within a time step Δt. Generally, employing smaller time steps yields more realistic results. Additionally, cells obstructed by natural features (e.g., rivers, rocks, roads) have a burning state of 0, and their burning speed is set to 0.

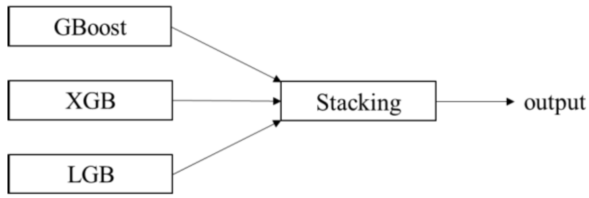

FIGURE 4. Stacking structure diagram in the burned area prediction model.

B. FFSRP MODEL

1) STACKING AND LEARNER SELECTION

ML allows for constructing statistical models by providing sufficient sample data and employing a series of algorithms to carry out predictions. It is found that ML models are advantageous in high prediction accuracy and flexible model structure [47]. In recent decades, the ensemble learning has attracted much attention in the field of ML because it can efficiently solve practical application problems () [28]. The basic process is: 1) generate a series of different learners; 2) use a certain integration method to combine learners to improve the prediction ability and generalization ability of the model. Boosting, Bagging, Stacking, and Blending are the most commonly used methods in ensemble learning [48]. Among them, Boosting and Bagging use similar learners for integration. Stacking and Blending are hierarchical structures, which can realize the integration of heterogeneous learners. Moreover, stacking uses K-fold cross validation, and learners are trained with all data. When the data set is not large, the performance of Stacking is more robust compared with Blending. Therefore, this paper uses Stacking to integrate multiple learners to establish the burned area of forest fire prediction model.

Stacking is the abbreviation of stacked generalization proposed by Wolpert in 1992. Its core idea is to carry out cross validation training on the base learner, form a training set based on the output of the base learner to train the meta learner, and the prediction result is formed by the output of the meta learner [49]. The base learner and meta learner are the core of the Stacking model, and the selection and combination of learners are the key to the integration of the Stacking model [50]. If the multi-layer stacking learner framework is adopted, it is very difficult to determine the best stacking method due to the large number of learners available and combinations. And the performance improvement after the integration of multi-layer stacking models is limited. Therefore, this paper selects a two-layer learner framework based on Stacking, chooses appropriate learners and their combinations, and constructs a prediction model to improve the prediction ability and generalization ability of the model.

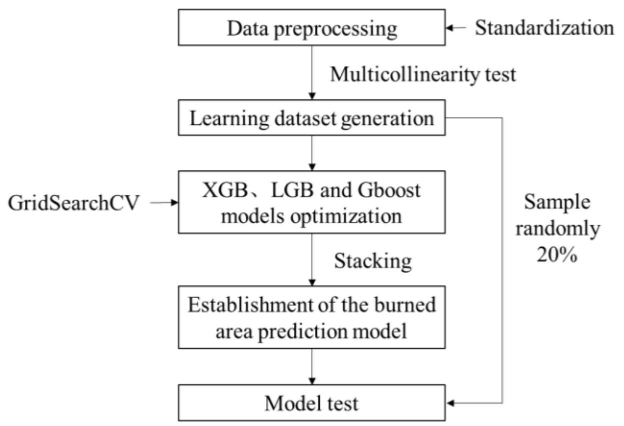

FIGURE 5. Flow chart for establishment of the burned area prediction model.

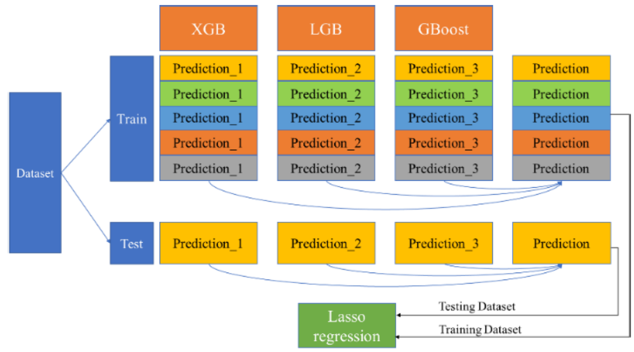

FIGURE 6. The training and testing process of the model.

To obtain a superior combination to all members, the selection of base learners should follow the principle of accuracy and diversity [51]. Common base learners with high accuracy include GBDT (GBoost), XGBoost (XGB), LightGBM (LGB), Random Forest and ANN which have the high performance in predicting forest fires [52], [53]. Especially, selecting the best forecasting model is a constant gamble because each ML algorithm has advantages and disadvantages [54]. It is necessary to choose the corresponding ML algorithm as much as possible based on other requirements under the premise of meeting the accuracy of prediction results. The burned area of forest fire prediction model needs to quickly process a large number of diverse types of data. Among these learners, GBoost can flexibly handle various types of data, including continuous and discrete values; XGB can realize the parallel operation of trees, greatly improving the speed of algorithm training and prediction [55]; LGB not only occupies low memory, but also has the ability to process big data. Therefore, this paper selects these three learners as the base learners of Stacking. Meta learner of Stacking should preferably be a simple model, such as Ridge regression, Lasso regression, to prevent over fitting of the overall model. Ridge and Lasso regressions can identify unimportant variables in the model and simplify the model. Compared with Ridge regression, Lasso regression can reduce some regression coefficients to eliminate variables. Therefore, this paper selects Lasso regression as the meta learner of Stacking. The structure of ensemble learning in this paper is shown in Fig. 4.

2) BURNED AREA PREDICTION ESTABLISHMENT PROCESS

The establishment process model is shown in Fig. 5. First of all, preprocess the historical data of forest fire area in the selected area (i.e., standardize it). Then, perform multicollinearity tests on the data, adjust variables, and make a learning dataset. Next, use GridsearchCV method to optimize XGB, LGB, and GBoost learners and establish the burned area prediction model based on Stacking to learn the "relationship" between the burned area and influencing factors on the training dataset. Finally, randomly select 20% of the samples from the learning dataset for model testing. Compare the predicted results with the actual results and investigate the reasons for abnormal samples.

3) MULTICOLLINEARITY TEST

Multicollinearity refers to the close correlation between explanatory variables in a linear regression model. Multicollinearity is common, and in general, moderate collinearity has little effect. However, severe collinearity can cause the significance test of explanatory variables to lose meaning and the model estimation to deviate or even be invalid. Therefore, when the model involves multiple explanatory variables, multicollinearity testing should be performed. This article uses the variance inflation factor (VIF) diagnostic method to perform multicollinearity tests on explanatory variables. Generally, the larger the VIF value, the more significant the multicollinearity between explanatory variables. It is generally believed that when the VIF is greater than 10, there is significant multicollinearity between explanatory variables.

4) MODEL OPTIMIZATION

In the ML models, the parameters that need to be manually selected are called hyperparameters. Improper selection of hyperparameters will lead to under fitting or over fitting. When selecting hyperparameters, there are two ways, one is to fine tune by experience, and the other is to select parameters of different sizes and bring them into the model to select the best performance parameters. One method of fine tuning is to manually modulate the hyperparameters until a good hyperparameter combination is found. This method is very time-consuming, so we use GridSearchCV of Scikit-Learn to do this search. In grid search, the parameters are searched, that is, within the specified parameter range, the parameters are adjusted in sequence by steps, and the adjusted parameters are used to train the learner to find the parameters with the highest accuracy in the verification dataset from all the parameters. This is actually a process of training and comparison. GridSearchCV can ensure to find the most accurate parameter within the specified parameter range. Therefore, based on Anaconda ML platform, this paper establishes

TABLE 5. Meteorological data during the natural development period of "3.29 forest fire".

Time	Wind speed (m/s)	Wind direction
17:00 on March 29th -20:00 on March 29th	4.75	west
20:00 on March 29th -02:00 on March 30th	4.00	west
02:00 on March 30th -08:00 on March 30th	4.00	southwest
08:00 on March 30th -09:00 on March 30th	3.34	southwest

TABLE 6. The propagation speed in each direction at each moment.

time	spread direction	R (m/min)	time	spread direction	R (m/min)
17:00 on March 29th - 20:00 on March 29th	OA	2.44	02:00 on March 30th - 08:00 on March 30th	OA	1.34
	OB	1.34		OB	0.80
	OC	0.74		OC	0.66
	OD	0.58		OD	0.80
	OE	0.74		OE	1.34
	OF	1.34		OF	2.23
	OG	2.44		OG	2.74
	OH	3.13		OH	2.23
20:00 on March 29th - 02:00 on March 30th	OA	2.23	08:00 on March 30th - 09:00 on March 30th	OA	1.34
	OB	1.34		OB	0.89
	OC	0.80		OC	0.74
	OD	0.66		OD	0.89
	OE	0.80		OE	1.34
	OF	1.34		OF	2.04
	OG	2.23		OG	2.43
	OH	2.74		OH	2.04

XGB, LGB, and GBoost learners, and uses GridSearchCV to optimize the model.

5) TRAINING AND TESTING OF THE MODEL

The training and testing process of the model is shown in Fig. 6, where XGB, LGB and GBoost algorithms are used as the base learners of Stacking [56]. To prevent over fitting of the model, a five-fold cross validation method is used to train the base learners. Taking XGB algorithm as an example, in the specific training process, the training set is divided into five equal copies. Four of them are used to train XGB, and the last one is predicted using the trained model. Then the XGB prediction results can be obtained with the base learners LGB and GBoost. The prediction results obtained by the base learners are used as new feature parameters and combined with the prediction target as a new training set. The meta learner Lasso regression model gives weight to the prediction results of the base learners through learning the training set, so that the prediction results are more accurate. During the model training, the error value is calculated by the following formula (8), where y_pred_i is the predicted burned area of the ith sample, y_true_i is the real burned area of the ith sample, and the calculation basis of the error value is the mean absolute error (MAE) between the predicted burned area and the real burned area.

$$Score = \frac{1}{m}\sum_{i=1}^{i=m} |y\_pred\_i - y\_true\_i| \tag{8}$$

Each model obtained by the base learner based on Stacking integration strategy through five-fold cross validation is used to test the model, and the average prediction results are used as the characteristic parameters for prediction, and output as the final prediction results.

IV. CASE STUDY

A. VERIFICATION OF FFSPP MODEL

1) DATA SOURCES

To validate the proposed methods in this study, the "3.29 Forest Fire" in southwestern China is taken as a case study [57]. Particularly, the "3.29 Forest Fire" in Anning was a major forest fire accident that occurred at 17:00 on March 29, 2006 in Wenquan Town, Anning City. The forest fire had a large burned area and a long duration, and its difficulty in extinguishing and danger coefficient were rare, reflecting the complexity, variability, and danger of typical mountain forest fires in the southwestern forest region of China. The fire began at 17:00 on March 29, 2006, and was completely extinguished on April 7, 2006. In these 10 days, the firefighting data were recorded by firefighters at 9:00 every day, and the daily range of changes in the fire boundary was also delineated. Due to entering the comprehensive firefighting stage after March 30th, the range of forest fire spread is no longer a natural development trend. Therefore, the calculation of forest fire spread in this study will be conducted until 9:00 on March 30th. The weather factors during this time period are shown in Table 5. Among them, the typical vegetation in the burned area is pinus yunnanensis. For the sake of simplification, the influence of slope on combustion speed is negligible in this work.

2) VALIDATION OF CA IMPROVEMENT

In addition, at 2:00 pm on March 29th, the temperature was 23.4°C, the humidity was 21%, and the wind speed was 10.00m/s (wind force level 5). According to formula (3)-(4), R_0 can be obtained as 4.83km/h. Select a cell edge length

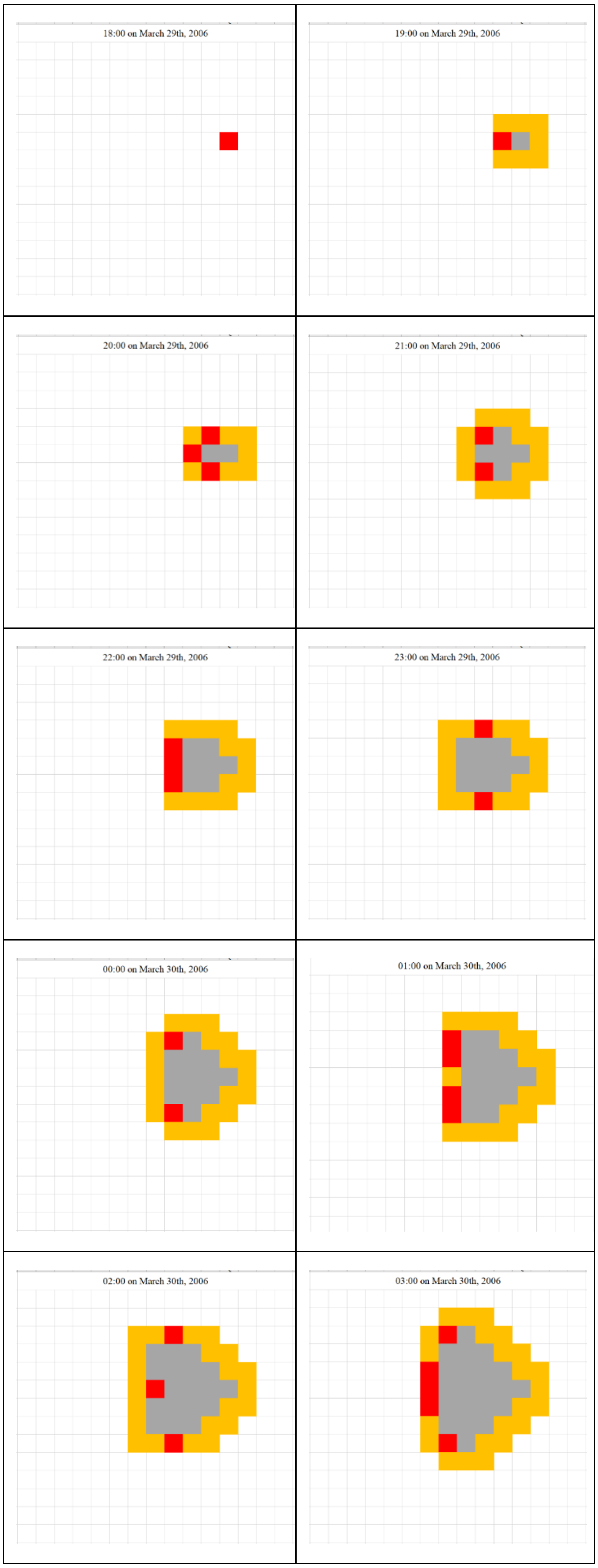

FIGURE 7. **Model simulation process of forest fire spread.**

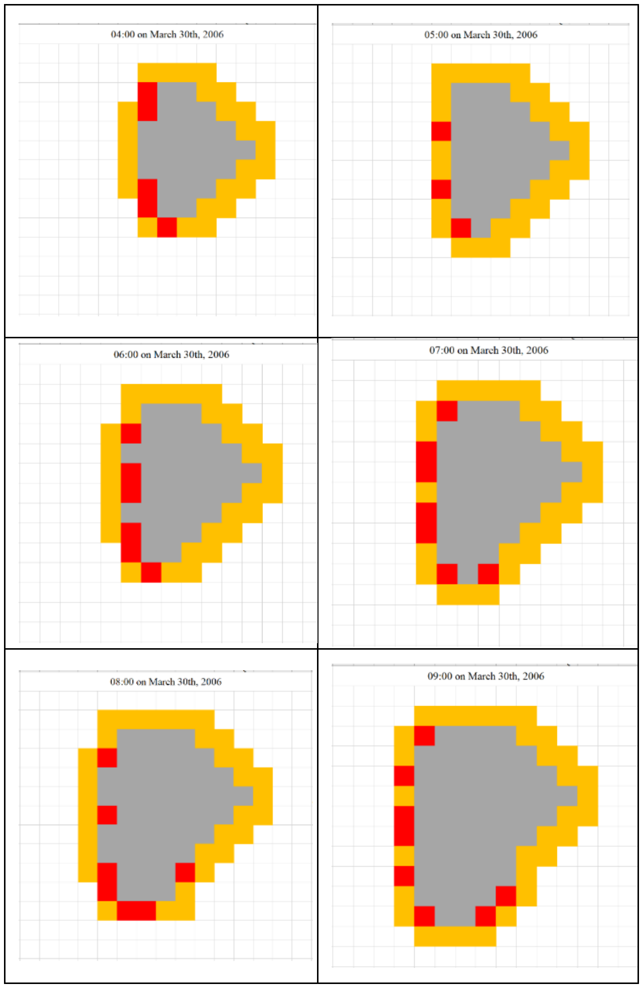

FIGURE 7. ***(Continued.)*** **Model simulation process of forest fire spread.**

of 1/6km and a time step of 1h for model calculation and simulation. The propagation speed in each direction at each moment is shown in Table 6.

The forest fire spread process simulated by the model is shown in Fig. 7, where red represents ''fully burned'', gold represents ''partially burned'', gray represents ''end burned'', and white represents ''unburned''.

The simulation results show that the burned area in the natural development period of the ''3.29 Forest Fire'' is 286.81 hm^2 (the actual fire area was 403.63 hm^2), with a relative error of 28.94%. The results indicate that the proposed model outperforms the Farsite and Prometheus fire behavior simulation models commonly used in US and Canadian industries. Among them, the Scott combustible model simulation result in Farsite is 939.66 hm^2 (relative error is 132.80%), the Anderson combustible model simulation result is 1089.19 hm^2 (relative error is 169.85%), and the FBP combustible model simulation result in Prometheus is 1587.20 hm^2 (relative error is 293.23%).

TABLE 7. Attribute information.

Number	Symbol	Meaning
1	X	x-axis spatial coordinate within the Montesinho park map: 1 to 9
2	Y	y-axis spatial coordinate within the Montesinho park map: 2 to 9
3	month	month of the year: 'jan' to 'dec'
4	day	day of the week: 'mon' to 'sun'
5	FFMC	FFMC index from the FWI system: 18.7 to 96.20
6	DMC	DMC index from the FWI system: 1.1 to 291.3
7	DC	DC index from the FWI system: 7.9 to 860.6
8	ISI	ISI index from the FWI system: 0.0 to 56.10
9	temp	temperature in Celsius degrees: 2.2 to 33.30
10	RH	relative humidity in %: 15.0 to 100
11	wind	wind speed in km/h: 0.40 to 9.40
12	rain	outside rain in mm/m^2 : 0.0 to 6.4
13	area	the burned area of the forest (in ha): 0.00 to 1090.84

TABLE 8. Mean and STD of data.

Data	Mean	STD	Data	Mean	STD
X	4.67	2.31	Y	4.30	1.23
month	7.48	2.28	Day	4.26	2.07
FFMC	90.64	5.52	DMC	110.87	64.05
DC	547.94	248.07	ISI	9.02	4.56
temp	18.89	5.81	RH	44.29	16.32
wind	4.02	1.79	rain	0.02	0.30
area	12.85	63.66			

TABLE 9. Partial data standardization results.

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
1.01	0.57	-1.97	0.36	-0.81	-1.32	-1.83	-0.86	-1.84	0.41	1.5	-0.07	0
1.01	-0.24	1.11	-1.09	-0.01	-1.18	0.49	-0.51	-0.15	-0.69	-1.74	-0.07	30
1.87	3.82	-0.21	-1.09	-0.88	-0.98	-0.95	-1.12	-0.15	-0.14	-0.74	-0.07	0.36
-1.59	-0.24	0.67	-1.09	0.06	0.29	0.58	-0.44	0.48	-0.39	-1.01	-0.07	0.43
-1.15	0.57	0.67	-1.57	0.05	0.24	0.56	-0.44	0.52	-0.32	-1.24	-0.07	0.47
-1.59	-1.87	0.23	-0.61	0.88	-0.17	-0.14	0.92	0.76	-0.81	0.27	-0.07	0.55
1.44	1.38	0.23	1.32	0.14	0.49	0.22	0.35	0.12	-0.2	0.99	-0.07	196.48
-1.15	-1.87	0.67	0.84	0.34	0.16	0.51	-0.09	-0.12	0.1	-1.24	-0.07	200.94
-1.59	-1.87	0.67	-1.09	0.06	0.29	0.58	-0.44	-0.02	-0.26	-1.01	-0.07	212.88
0.58	0.57	0.67	0.84	0.34	0.16	0.51	-0.09	1.07	-1.06	-0.01	-0.07	1090.84

B. VERIFICATION OF FFSRP MODEL

1) DATA PROCESSING

a: ACQUISITION OF DATA

This paper uses a real fire dataset from Montesinho National Forest Park in Portugal from January 2000 to December 2003.

TABLE 10. Multicollinearity test of variables.

variable	X	day	DMC	DC	ISI	RH	wind	rain
VIF value	4.50	4.82	7.96	9.48	5.32	7.15	5.40	1.03

TABLE 11. The initial values of each parameter in the XGB model.

Parameter	Initial value	Parameter	Initial value
n_estimators	500	max_depth	5
min_child_weight	1	gamma	0
subsample	0.8	colsample_bytree	0.8
reg_alpha	0	reg_lambda	1
learning_rate	0.1		

TABLE 12. The order, range, and step size of the adjustment parameters for the XGB model.

Order	Parameter	Range	Step size
1	n_estimators	[50,800]	1
2	max_depth	[1,10]	1
	min_child_weight	[1,6]	0.01
3	gamma	[0,0.4]	0.01
4	subsample	[0.7,1]	0.01
	colsample_bytree	[0.4,1]	0.1
5	reg_alpha	[0.1,3.5]	0.1
	reg_lambda	[0.5,3]	0.1
6	learning_rate	[0.01,0.2]	0.01

TABLE 13. The optimization results of each parameter in the XGB model.

Parameter	Optimization result	Parameter	Optimization result
n_estimators	50	max_depth	2
min_child_weight	1	gamma	0
subsample	1	colsample_bytree	0.9
reg_alpha	2.7	reg_lambda	1.8
learning_rate	0.01		

The dataset contains 13 variables and 517 entries. Among them, the four variables FFMC, DMC, DC, and ISI come from the FWI subsystem of the Canadian Forest Fire Risk Rating System. The names and meanings of each variable are shown in Table 7. First, the dates in the dataset are digitized and encoded, and then the digital features of the data are viewed. The burned area is designated as the result variable, and the other feature vectors are designated as the dependent variable. By examining the digital features of the burned area, it can be seen that the mean is 12.85, with the top 99.22% of data being less than 154, the top 75% of data being less than 6.57, and the top 53% of data being less than 1.01. It indicates that during this period, small and medium-sized forest fires mainly occurred in the region.

b: STANDARDIZATION OF DATA

Data standardization has evolved into a common method of data processing in ML [58]. It serves to mitigate model overfitting, eliminate bias against any influencing factors,

TABLE 14. Training results of various ML models.

Model	MAE	STD
XGB	22.77	4.75
LGB	20.91	4.00
GBoost	21.82	5.08
FFSRP model	**16.50**	**2.58**

TABLE 15. Partial predicted and actual burned area on the learning dataset.

number	true	predict
200	11.53	14.14
201	12.10	12.24
202	13.05	12.87
203	13.70	9.27
204	13.99	12.24
205	14.57	14.18
206	15.45	13.99
207	17.20	14.47

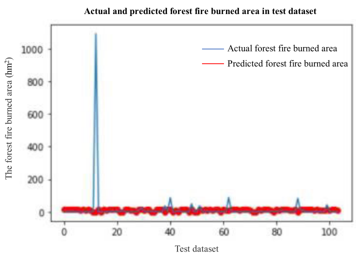

FIGURE 8. Actual and predicted forest fire burned area graph.

and enhance overall model accuracy. Simultaneously, data standardization addresses issues such as gradient explosion, gradient disappearance, and overfitting in the domains of deep learning and reinforcement learning. As shown in Table 8, since there is a significant difference between the mean and standard deviation (STD) of the factors affecting the evacuation time, this paper adopts formula (9) to standardize the data., where μ and σ are the mean and STD of the influencing factors corresponding to x_i. The processed partial data are shown in Table 9. To facilitate the comparison between the predicted and actual evacuation time, the exit time (s) is not standardized.

$$x_i = \frac{x\_i - \mu}{\sigma} \tag{9}$$

c: MULTICOLLINEARITY TEST

After inspection, the VIF of 8 influencing factors including X, day, DMC, DC, ISI, RH, wind, and rain is less than 10 (Table 10), indicating that the corresponding data can enter the model fitting stage.

2) OPTIMIZATION OF XGB, LGB, AND GBOOST MODELS

Employing GridSearchCV to optimize the XGB, LGB, and GBoost models, we illustrate the optimization process using the XGB model as an example. The XGBoost model features an extensive array of hyperparameters, originating from both the XGBoost algorithm and the decision tree. These hyperparameters play a direct role in shaping the overall accuracy and predictive performance of the algorithm. The identification of an optimal combination of hyperparameters can significantly enhance the XGBoost model performance. The detailed optimization process is outlined below.

a: SET INITIAL VALUE

The initial values of each parameter in the XGB model are set and preented in Table 11.

b: ADJUST PARAMETERS

The order, range, and step size of the adjustment parameters for the XGB model are shown in Table 12:

c: ADJUST PARAMETERS

The optimization results of each parameter in the XGB model are shown in Table 13:

3) MODEL TRAINING AND TESTING

In this study, FFSRP model is established by integrating XGB, LGB, and GBoost learners which are optimized by the network search method with stacking strategy based on the Anaconda ML platform. In addition to FFSRP model, three ML models - XGB, LGB and GBoost - were used to predict the burned area of each fire in Montesinho National Forest Park in Portugal from January 2000 to December 2003 (517 samples are set in the dataset). The calculation basis of the error value is the MAE between predicted and actual burned area. The training results of various ML models are shown in Table 14: the MAE and STD values of FFSRP model are both smaller than those of XGB, LGB and GBoost models. That is, the result of FFSRP model is better than XGB, LGB and GBoost models.

To test the generalization ability of the model, 103 samples (not trained) are randomly selected from the dataset for testing. Table 15 compares partial predicted and actual burned area on the learning dataset ("number" refers to the number of data in the dataset). As illustrated in Figure 8, this study discerns values exhibiting a substantial disparity between predicted and actual fire areas, revealing that the predictive accuracy for large fires is inferior compared to that for small and medium-sized fires. This discrepancy arises from the heightened complexity in the occurrence and development processes of large fires, rendering their consequences more challenging to predict. Consequently, the model excels in achieving robust predictive performance in scenarios involving small and medium-sized fires.

V. CONCLUSION

This paper presents a FFSRP model after an accidental forest fire occurrence. First, based on relevant literature and actual situations, the forest fire influencing factors that are easily obtainable during the practical application of the models have been determined: combustible state, meteorological factors, and topography conditions. Then, the FFSPP model is established by combining the CA model and Wang Zhengfei model. In addition, the FFSRP model is proposed to predict the burned area through ML methods. Finally, the "3.29 Forest Fire" in China and the real fire dataset from Montesinho

National Forest Park in Portuga are taken as the examples to verify the proposed model and method. It is found that, the relative error of the proposed FFSPP model is smaller than the relative error of the Farsite and Prometheus fire behavior simulation models. And the proposed FFSRP model can achieve good predictive performance in small and medium-sized fire conditions. The main research results are as follows:

(1) The factors influencing forest fire are determined. The main factors affecting forest fires are combustible state, meteorological factors, and topography conditions. Among them, combustible state mainly refers to plant type, meteorological factors include temperature, humidity, wind speed, wind direction, FWI, etc., topography conditions include geographical location, slope, etc., all of which are easy to obtain during the practical application of the models

(2) The proposed FFSPP model has the capability to describe the process of fire propagation. The FFSPP model is established by combining the CA model and Wang Zhengfei model to predict the direction and speed of forest fire spread and achieve visualization. The results show that the proposed model outperforms the Farsite and Prometheus fire behavior simulation models commonly used in the US and Canadian industries.

(3) The proposed FFSRP model is able to describe the fire effect. Based on the Anaconda ML platform, the FFSRP model is established by integrating XGB, LGB and GBoost learners with stacking. The final MAE value of the model on the training dataset is 16.50, which can achieve a good prediction effect especially for small and medium-sized fire situations.

Despite the above novelties and contributions, this work still has several limitations. Due to limited practical conditions, this study only used past practical cases for model verification and did not conduct combustion experiments on forests. In addition, the FFSBP model proposed in this paper still has a lot of room for improvement with limited samples. Therefore, potential future research works can be done to: 1) set aside experimental areas for forest burning experiments; 2) expand the sample to improve the accuracy of the forest fire spread prediction model.

ACKNOWLEDGMENT

This work was supported by Interdisciplinary Research Project for Young Teachers of USTB (Fundamental Research Funds for the Central Universities) No. FRF-IDRY-21-016 and Researchers Supporting Project number (RSP2024R388), King Saud University, Riyadh, Saudi Arabia.

REFERENCES

[1] M. Flannigan, B. Stocks, M. Turetsky, and M. Wotton, "Impacts of climate change on fire activity and fire management in the circumboreal forest," *Global Change Biol.*, vol. 15, no. 3, pp. 549–560, Mar. 2009.

[2] M. R. Turetsky, E. S. Kane, J. W. Harden, R. D. Ottmar, K. L. Manies, E. Hoy, and E. S. Kasischke, "Recent acceleration of biomass burning and carbon losses in Alaskan forests and peatlands," *Nature Geosci.*, vol. 4, no. 1, pp. 27–31, Jan. 2011.

[3] N. Ntinopoulos, M. Spiliotopoulos, L. Vasiliades, and N. Mylopoulos, "Contribution to the study of forest fires in semi-arid regions with the use of Canadian fire weather index application in Greece," *Climate*, vol. 10, no. 10, p. 143, Sep. 2022.

[4] X. Deng, Z. Zhang, F. Zhao, Z. Zhu, and Q. Wang, "Evaluation of the regional climate model for the forest area of Yunnan in China," *Frontiers Forests Global Change*, vol. 5, Jan. 2023, Art. no. 1073554.

[5] Y. Pang, Y. Li, Z. Feng, Z. Feng, Z. Zhao, S. Chen, and H. Zhang, "Forest fire occurrence prediction in China based on machine learning methods," *Remote Sens.*, vol. 14, no. 21, p. 5546, Nov. 2022.

[6] C. Gao, H. Lin, and H. Hu, "Forest-fire-risk prediction based on random forest and backpropagation neural network of Heihe area in Heilongjiang province, China," *Forests*, vol. 14, no. 2, p. 170, Jan. 2023.

[7] C. Shi and F. Zhang, "A forest fire susceptibility modeling approach based on integration machine learning algorithm," *Forests*, vol. 14, no. 7, p. 1506, Jul. 2023.

[8] T. Artés, A. Cencerrado, A. Cortés, and T. Margalef, "Relieving the effects of uncertainty in forest fire spread prediction by hybrid MPI-OpenMP parallel strategies," *Proc. Comput. Sci.*, vol. 18, pp. 2278–2287, Jan. 2013.

[9] D. W. Xie and S. L. Shi, "Prediction for burned area of forest fires based on SVM model," *Appl. Mech. Mater.*, vols. 513–517, pp. 4084–4089, Feb. 2014.

[10] J. N. S. Rubí, P. H. P. de Carvalho, and P. R. L. Gondim, "Application of machine learning models in the behavioral study of forest fires in the Brazilian federal district region," *Eng. Appl. Artif. Intell.*, vol. 118, Feb. 2023, Art. no. 105649.

[11] M. Denham, K. Wendt, G. Bianchini, A. Cortés, and T. Margalef, "Dynamic data-driven genetic algorithm for forest fire spread prediction," *J. Comput. Sci.*, vol. 3, no. 5, pp. 398–404, Sep. 2012.

[12] V. G. Ntinas, B. E. Moutafis, G. A. Trunfio, and G. C. Sirakoulis, "Parallel fuzzy cellular automata for data-driven simulation of wildfire spreading," *J. Comput. Sci.*, vol. 21, pp. 469–485, Jul. 2017.

[13] L. Sun, C. Xu, Y. He, Y. Zhao, Y. Xu, X. Rui, and H. Xu, "Adaptive forest fire spread simulation algorithm based on cellular automata," *Forests*, vol. 12, no. 11, p. 1431, Oct. 2021.

[14] X. Xin, H. Jiang, G. Zhou, S. Yu, and Y. Wang, "Canadian forest fire weather index (FWI) system: A review," *J. Zhejiang A&F Univ.*, vol. 28, no. 2, pp. 314–318, 2011.

[15] C. Li, L. Zhao, D. Li, and J. Ma, "Prediction model of forest fire extension in Australia and application of emergency materials dispatching based on CA system," *Forest Inventory Planning*, vol. 45, no. 5, pp. 62–69, 2020.

[16] A. Karouni, B. Daya, S. Bahlak, and P. Chauvet, "A simplified mathematical model for fire spread predictions in wildland fires combining between the models of Anderson and Rothermel," *Int. J. Model. Optim.*, vol. 4, no. 3, pp. 197–200, Jun. 2014.

[17] C. Li, Z. Jing, B. Tang, and Y. Zhang, "Analysis of forest fire spread trend surrounding transmission line based on Rothermel model and Huygens principle," *Int. J. Multimedia Ubiquitous Eng.*, vol. 9, no. 9, pp. 51–60, Sep. 2014.

[18] X. Zhang, P. Liu, and X. Wang, "Research on improvement of Wang Zhengfei's forest fire spread model," *J. Shandong Forestry Sci. Technol.*, vol. 50, no. 1, pp. 1–6, 2020.

[19] A. R. Ervilha, J. M. C. Pereira, and J. C. F. Pereira, "On the parametric uncertainty quantification of the Rothermel's rate of spread model," *Appl. Math. Model.*, vol. 41, pp. 37–53, Jan. 2017.

[20] Y. Zhang and L. Tian, "Examining and reforming the Rothermel surface fire spread model under no-wind and zero-slope conditions for the Karst ecosystems," *Forests*, vol. 14, no. 6, p. 1088, May 2023.

[21] M. Feč kan and J. Pač uta, "Existence of solution of a forest fire spread model," *Appl. Math. Lett.*, vol. 83, pp. 227–231, Sep. 2018.

[22] U. Kc, J. Aryal, S. Garg, and J. Hilton, "Global sensitivity analysis for uncertainty quantification in fire spread models," *Environ. Model. Softw.*, vol. 143, Sep. 2021, Art. no. 105110.

[23] Z. Zheng, W. Huang, S. Li, and Y. Zeng, "Forest fire spread simulating model using cellular automaton with extreme learning machine," *Ecol. Model.*, vol. 348, pp. 33–43, Mar. 2017.

[24] X. Rui, S. Hui, X. Yu, G. Zhang, and B. Wu, "Forest fire spread simulation algorithm based on cellular automata," *Natural Hazards*, vol. 91, no. 1, pp. 309–319, Mar. 2018.

[25] M. H. Vahidnia, A. A. Alesheikh, S. Behzadi, and S. Salehi, "Modeling the spread of spatio-temporal phenomena through the incorporation of ANFIS and genetically controlled cellular automata: A case study on forest fire," *Int. J. Digit. Earth*, vol. 6, no. 1, pp. 51–75, Jan. 2013.

[26] T. Yu, H. Ding, W. Huang, and J. Qiao, "A survey of ensemble learning for complex heterogeneous data," *Control Eng. China*, vol. 30, no. 8, pp. 1425–1435, 2023.

[27] C.-X. Zhang and J.-S. Zhang, "A survey of selective ensemble learning algorithms," *Chin. J. Comput.*, vol. 34, no. 8, pp. 1399–1410, Oct. 2011.

[28] J. Xu and Y. Yang, "A survey of ensemble learning approaches," *J. Yunnan Univ. Natural Sci. Ed.*, vol. 40, no. 6, pp. 1082–1092, 2018.

[29] Z. Wu, M. Li, B. Wang, Y. Tian, Y. Quan, and J. Liu, "Analysis of factors related to forest fires in different forest ecosystems in China," *Forests*, vol. 13, no. 7, p. 1021, Jun. 2022.

[30] F. Guo, S. Selvalakshmi, F. Lin, G. Wang, W. Wang, Z. Su, and A. Liu, "Geospatial information on geographical and human factors improved anthropogenic fire occurrence modeling in the Chinese boreal forest," *Can. J. Forest Res.*, vol. 46, no. 4, pp. 582–594, Apr. 2016.

[31] H. K. Preisler and A. L. Westerling, "Statistical model for forecasting monthly large wildfire events in Western United States," *J. Appl. Meteorol. Climatol.*, vol. 46, no. 7, pp. 1020–1030, Jul. 2007.

[32] A. C. Zeng, Q. J. Cai, Z. Su, X. B. Guo, Q. F. Jin, and F. T. Guo, "Seasonal variation and driving factors of forest fire in Zhejiang Province, China, based on MODIS satellite hot spots," *J. Appl. Ecol.*, vol. 31, no. 2, pp. 399–406, 2020.

[33] W. Ma, Z. Feng, Z. Cheng, S. Chen, and F. Wang, "Identifying forest fire driving factors and related impacts in China using random forest algorithm," *Forests*, vol. 11, no. 5, p. 507, May 2020.

[34] J. Li, Y. Shan, S. Yin, M. Wang, L. Sun, and D. Wang, "Nonparametric multivariate analysis of variance for affecting factors on the extent of forest fire damage in Jilin province, China," *J. Forestry Res.*, vol. 30, no. 6, pp. 2185–2197, Dec. 2019.

[35] S. A. Perestrelo, M. C. Grácio, N. D. A. Ribeiro, and L. M. Lopes, "A multi-scale network with percolation model to describe the spreading of forest fires," *Mathematics*, vol. 10, no. 4, p. 588, Feb. 2022.

[36] S. Hui, X. Rui, and Y. Li, "An lmproved forest fire spread simulation algorithm coupled with cellular automata," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 41, no. 10, pp. 1326–1332, 2016.

[37] N. M. Gharakhanlou and N. Hooshangi, "Dynamic simulation of fire propagation in forests and rangelands using a GIS-based cellular automata model," *Int. J. Wildland Fire*, vol. 30, no. 9, pp. 652–663, Jun. 2021.

[38] O. Jellouli, A. Bernoussi, M. Mâatouk, and M. Amharref, "Forest fire modelling using cellular automata: Application to the watershed oued laou (Morocco)," *Math. Comput. Model. Dyn. Syst.*, vol. 22, no. 5, pp. 493–507, Sep. 2016.

[39] X. Li, M. Zhang, S. Zhang, J. Liu, S. Sun, T. Hu, and L. Sun, "Simulating forest fire spread with cellular automation driven by a LSTM based speed model," *Fire*, vol. 5, no. 1, p. 13, Jan. 2022.

[40] Y. Xu, D. Li, H. Ma, R. Lin, and F. Zhang, "Modeling forest fire spread using machine learning-based cellular automata in a GIS environment," *Forests*, vol. 13, no. 12, p. 1974, Nov. 2022.

[41] F. Allaire, V. Mallet, and J.-B. Filippi, "Emulation of wildland fire spread simulation using deep learning," *Neural Netw.*, vol. 141, pp. 184–198, Sep. 2021.

[42] P. Cortez and A. Morais, "A data mining approach to predict forest fires using meteorological data," Associação Portuguesa para a Inteligência Artif., Portuguese Republic, Tech. Rep., pp. 512–523, 2007.

[43] M. Bisquert, E. Caselles, J. M. Sánchez, and V. Caselles, "Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data," *Int. J. Wildland Fire*, vol. 21, no. 8, p. 1025, 2012.

[44] T. G. Rad and M. Karimi, "Evaluation performances of different forest fire spread models using cellular automata case study The forests of Lakan district in Rasht," *xxxx*, vol. 23, no. 1, pp. 64–78, 2015.

[45] T. Sun, L. Zhang, W. Chen, X. Tang, and Q. Qin, "Mountains forest fire spread simulator based on geo-cellular automaton combined with Wang Zhengfei velocity model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 4, pp. 1971–1987, Aug. 2013.

[46] M. Xianmin, "The influence of wind and relief on the speed of the forest fire spreading," *J. Appl. Meteorol. Sci.*, vol. 4, no. 1, pp. 100–104, 1993.

[47] O. Avalos, "GSA for machine learning problems: A comprehensive overview," *Appl. Math. Model.*, vol. 92, pp. 261–280, Apr. 2021.

[48] J. Shi, T. Yu, K. Goebel, and D. Wu, "Remaining useful life prediction of bearings using ensemble learning: The impact of diversity in base learners and features," *J. Comput. Inf. Sci. Eng.*, vol. 21, no. 2, Apr. 2021.

[49] Z. Zhou, "Machine learning," *China Non-Governmental*, vol. 3, no. 21, pp. 171–196, 2016.

[50] P. Melville and R. J. Mooney, "Creating diversity in ensembles using artificial data," *Inf. Fusion*, vol. 6, no. 1, pp. 99–111, Mar. 2005.

[51] M. P. Sesmero, A. I. Ledezma, and A. Sanchis, "Generating ensembles of heterogeneous classifiers using stacked generalization," *WIREs Data Mining Knowl. Discovery*, vol. 5, no. 1, pp. 21–34, Jan. 2015.

[52] Z. Wang, K. Wang, Y. Li, and G. Li, "Research on forest fire prediction in Yunnan province based on LightGBM and SHAP," *Fire Science Technol.*, vol. 42, no. 11, pp. 1567–1571, 2023.

[53] M. Seddouki, M. Benayad, Z. Aamir, M. Tahiri, M. Maanan, and H. Rhinane, "Using machine learning coupled with remote sensing for forest fire susceptibility mapping. Case study Tetouan province, Northern Morocco," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLVIII-4/W6, pp. 333–342, Feb. 2023.

[54] R. Alkhatib, W. Sahwan, A. Alkhatieb, and B. Schütt, "A brief review of machine learning algorithms in forest fires science," *Appl. Sci.*, vol. 13, no. 14, p. 8275, Jul. 2023.

[55] J. Shi and J. Zhang, "Load forecasting based on multi-model by stacking ensemble learning," *Proc. CSEE*, vol. 39, no. 14, pp. 4032–4042, 2019.

[56] Z. Zhou, *Integrated Learning: Fundamentals and Algorithms*. Beijing, China: Publishing House of Electronics Industry, 2020.

[57] Z. Fan, L. Shu, R. Zhou, X. Xiao, M. Wang, F. Zhao, and Q. Wang, "Evaluation of forest fire behavior simulation model in southwest forest region," *Chin. J. Appl. Ecol.*, vol. 28, no. 10, pp. 3144–3154, 2017.

[58] M. K. Wes, *Python for Data Analysis*. Sevastopol, CA, USA: O'Reilly Media, 2013.

XUAN SUN was born in June 1979. He received the Ph.D. degree from the Public Security Research Institute, Tsinghua University. He was a First-Class Registered Fire Engineer and a Registered Safety Engineer. He is currently a Researcher. He is also the Deputy Director of the Disaster Prevention Research Center of the Ministry of Housing and Urban-Rural Development, the Director of the Institute of Building Fire Protection, China Academy of Building Research, the Director of the Building Fire Protection Professional Committee of the China Engineering Construction Standardization Association, the Director of the Building Fire Protection Professional Committee of the China Fire Protection Association, the Director of the Public Security Research Center, China Academy of Building Research, a member of the National Public Security Basic Standardization Technical Committee, a member of the National Fire Standardization Committee, and a member of the Expert Committee of the Disaster Prevention Research Center of the Ministry of Housing and Urban-Rural Development. He has undertaken the research work of more than 20 national "11th Five-Year" Science and Technology Support Plan Projects, National "12th Five-Year" Science and Technology Support Plan Projects, Major Scientific Research Projects, and Related Projects of Beijing Municipal Commission of Science and Technology. He has participated in the compilation of six books, including *Introduction to Public Safety Science*. He has participated in the compilation of national standards, such as fire protection code for garage, repair garage and parking lot design, and building fire load code.

NING LI was born in Henan, China, in 1997. She received the B.S. degree in safety engineering and the M.S. degree in safety science and engineering from the University of Science and Technology Beijing, in 2020 and 2023, respectively. She is currently a full-time Teacher with Xinjiang Institute of Technology. Her research interests include predicting dynamic evacuation of personnel and predicting the spread of forest fires. She has participated in the National Key Research and Development Program of China and published two articles and two invention patents. Her awards and honors include the Merit Graduates and the Second-Class Scholarship for postgraduate students from the University of Science and Technology Beijing.

DUOQI CHEN was born in 1983. He is currently the Deputy Captain and a Senior Engineer of the Fire Rescue Detachment in Tongzhou, Beijing. In the aspect of academic research, as the first editor, he has published the book *Research on the Theory and Practice of Fire Protection Supervision and Management* and published *Analysis of Fire Accidents of Death and Injury in Densely Populated Places* in the national core journal *Journal of Guangxi University*. He has 16 related papers have been published in other professional journals with profound academic foundation and rich theoretical achievements. In terms of practical experience, he has held different positions, such as fire control supervision, administrative license and key service, and with rich practical experience and excellent organizational ability. He has organized and completed the construction, examination and acceptance of key projects, such as Beijing Deputy Central Administrative Office District, Beijing Global Resort and Canal Business District, and made important contributions in the field of fire control. He is currently a member of the Communist Party of China (CPC).

GUANG CHEN was born in September 1982. He received the master's degree in safety engineering from Tsinghua University. Since 2005, he has been engaged in fire safety work for a long time. He has rich working experience and theoretical foundation. He owns two patents and five software copyrights. He participated in the compilation of one book and the formulation of two local standards, published academic papers in periodicals for many times, and organized three information platforms in the field of fire safety. In August 2023, the Development and Application of the Cloud Platform for Joint Prevention and Control for Museum architecture Fires, developed as a member of the leading team, and won the Second Prize of Science and Technology Innovation of the 13th China Fire Protection Association.

CHANGJUN SUN was born in Dalian, Liaoning, in April 1982. He received the bachelor's degree in fire engineering from China Armed Police Force College, in June 2013. He has published five papers, invention software works one (software name: command building fire safety real-time monitoring management system V1.0, certificate number: soft deng word 8744929), since 2022, the system used in Beijing 3.3 building, fully provide equipment operation of digital early warning, intelligent diagnosis, gradually improve the support by information means, and with resource sharing shared share the wisdom of the regulation mode. Realize the comprehensive, remote, and all-weather real-time monitoring of firefighting facilities by the client. Run data cloud storage and archiving, to achieve data easy traceability, hidden trouble record, and the ability of accountability. Provide reliable technical support for management and rescue. He is currently a member of the Communist Party of China.

MULIN SHI is currently the Deputy Director of Binzhou Safety Science and Technology Service Center, a Senior Accountant, a National Campus Safety and Health Expert, a Shandong Provincial Government Procurement Evaluation Expert, a Binzhou Municipal Government Procurement Evaluation Expert, and a Shandong Provincial Commerce and Trade Industry Enterprise Safety Production Standardization Evaluation Expert. Over the years, combined with the practical problems encountered in the work, she has published a number of articles to explain the point of view, and applied in the work. Among them, the article "Try to analyze the causes of distortion of accounting information and countermeasures,"s which provides practical measures for the macro-control of relevant government departments, was awarded the third prize of the 21st Outstanding Achievements in Social Sciences of Binzhou City. In recent years, she has led the technical team to engage in the technical support service and research work of departmental emergency management. In 2020, she presided over the technical support of the Binzhou pilot project of the national risk and disaster survey. In 2022, she presided over the normal dynamic update application tracking pilot project (Binzhou) project of the census and evaluation of the National Disaster Reduction Committee.

XUEHONG GAO was born in 1990. He received the Ph.D. degree in industrial engineering from Pusan National University, South Korea. He is currently an Associate Professor Faculty Member with the University of Science and Technology Beijing. China. He has published more than 30 research articles in journals, such as *IJPE*, *JCLP*, *IJPR*, *CAIE*, and *ANOR*. His research interests include robust optimization, disaster management, safety science, forest fires, and operations research.

KUO WANG was born in 2001. He received the bachelor's degree in engineering, with a major in safety engineering from Hebei University of Technology, in 2022. He is currently pursuing the master's degree with the University of Science and Technology Beijing. He is also engaged in research on forest fires, he has published one SCI article in the *Journal of Building Engineering*. His master's research is focused on the recognition of forest fires using unmanned aerial vehicles (UAVs) and distance measurement.

IBRAHIM M. HEZAM received the Ph.D. degree in operations research and decision support from Menoufia University, Egypt. He held a Post-Doctoral position in industrial engineering at Pusan National University, Busan, South Korea. He is currently an Associate Professor of operations research with the king Saud University, KSA. His research interests are artificial intelligence, optimization, sustainability, operations research, and decision support.