

Article

Estimation of Frost Hazard for Tea Tree in Zhejiang Province Based on Machine Learning

Jie Xu ^{1,2,3}, Suri Guga ^{1,2,3}, Guangzhi Rong ^{1,2,3}, Dao Riao ^{1,2,3}, Xingpeng Liu ^{1,2,3}, Kaiwei Li ^{1,2,3} and Jiquan Zhang ^{1,2,3,*}

¹ School of Environment, Northeast Normal University, Changchun 130024, China; xuj463@nenu.edu.cn (J.X.); surgg146@nenu.edu.cn (S.G.); ronggz728@nenu.edu.cn (G.R.); daora388@nenu.edu.cn (D.R.); liuxp912@nenu.edu.cn (X.L.); likw395@nenu.edu.cn (K.L.)

² Department of Environment, Institute of Natural Disaster Research, Northeast Normal University, Changchun 130024, China

³ Key Laboratory for Vegetation Ecology, Ministry of Education, Changchun 130024, China

* Correspondence: zhangjq022@nenu.edu.cn

Abstract: Tea trees are the main economic crop in Zhejiang Province. However, spring cold is a frequent occurrence there, causing frost damage to the valuable tea buds. To address this, a regional frost-hazard early-warning system is needed. In this study, frost damage area was estimated based on topography and meteorology, as well as longitude and latitude. Based on support vector machine (SVM) and artificial neural networks (ANNs), a multi-class classification model was proposed to estimate occurrence of regional frost disasters using tea frost cases from 2017. Results of the two models were compared, and optimal parameters were adjusted through multiple iterations. The highest accuracies of the two models were 83.8% and 75%, average accuracies were 79.3% and 71.3%, and Kappa coefficients were 79.1% and 67.37%. The SVM model was selected to establish spatial distribution of spring frost damage to tea trees in Zhejiang Province in 2016. Pearson's correlation coefficient between prediction results and meteorological yield was 0.79 ($p < 0.01$), indicating consistency. Finally, the importance of model factors was assessed using sensitivity analysis. Results show that relative humidity and wind speed are key factors influencing accuracy of predictions. This study supports decision-making for hazard prediction and defense for tea trees facing frost.



Citation: Xu, J.; Guga, S.; Rong, G.; Riao, D.; Liu, X.; Li, K.; Zhang, J. Estimation of Frost Hazard for Tea Tree in Zhejiang Province Based on Machine Learning. *Agriculture* **2021**, *11*, 607. <https://doi.org/10.3390/agriculture11070607>

Academic Editors:

Selma Boumerdassi, Eric Renault, Christopher Robin Bryant and Francesco Marinello

Received: 17 May 2021

Accepted: 24 June 2021

Published: 29 June 2021

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



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: tea tree; frost disaster; machine learning; frost hazard; space distribution

1. Introduction

Tea is a traditional drink, with a history that can be traced back 5000 years, and which has profound cultural and economic significance [1]. Tea plants are a type of warm-leaf plant. The shrub-type tea plants in the middle and lower reaches of the Yangtze River in China maintain a good growth state in 25–30 °C, and the sprout temperature of tea plants is 6–12 °C [2,3]. As the temperature rises in the early spring, the cold-resistant ability of tea trees decreases after the sprouting of tea buds, which can be damaged by freezing if the temperature drops sharply to below 0 °C. Frost disasters cause destruction of the tea protoplasm when the water in tea-leaf cells freezes, and this reduces tea yield [4]. Frost damage to the tea bud not only affects the quality and taste of tea, but also stops the germination of tea buds, causes bud death, and delays the picking period for spring tea [5,6].

Frost is a type of agricultural meteorological disaster. Frost disasters are caused by a strong cold wave in the crop-growing season, where the temperature of plants and leaves drops to below 0 °C and growing plants suffer from frost damage, leading to reduced crop yield, crop failure, or quality decline [7]. Frost disasters can be caused by two weather processes, radiation and advection; frost disasters caused by radiation are more common in Zhejiang [8–10]. Radiation frost disasters occur with a decrease in apparent heat, caused

by the loss of net energy from the surface to the sky by radiation under conditions of clear skies and very little wind [11]. Climate change frequently causes climate fluctuation events, resulting in an increased probability of frost disaster events [12]. Because of the increasing frost caused by the uncertainty of climate change, low temperatures and frost threat are increasing in the spring tea-planting areas, making them sensitive to climate change.

The influence of spring frost disasters is widespread and serious. To establish the impact of large-scale spring frost damage quickly and effectively, the normalized difference vegetation index (NDVI), normalized NDVI valley area index (NNVAI), spring frost damage index (SFDI), and other indices have been proposed and calculated; these are based on remote-sensing images that can be used to conduct real-time assessment of the damage to crops caused by spring frost [13–15]. Several studies have analyzed the duration and severity of frost events using historical meteorological data; the distribution characteristics of frost damage were analyzed, and the recurrence period of frost damage was calculated to strengthen the management of frost risk [16]. In small-scale areas, researchers have focused mainly on meteorological factors, which cannot reflect the detailed characteristics of frost damage [17]. When studying frost in mountainous areas, the main control parameter of temperature distribution in the complex terrain is altitude. Generally, the decreasing rate of air temperature is set at 0.65 °C/100 m [18,19], but it also fluctuates due to long-wave radiation and other factors, and even leads to temperature inversion. On a small scale, the slope aspect and the curvature of the terrain affect solar emissivity and local circulation, resulting in a difference in low-temperature distribution [20–22]. In several studies, surface temperature data was obtained by satellite remote-sensing and coupled with terrain factors (e.g., aspect, slope) to establish a low-temperature model that accurately reflected the spatial distribution of frost [23–25]. Researchers have applied several methods to model frost events in complex terrain, including multi-variate adaptive regression splines (MARS) [26], logistic regression and decision trees [27], and fuzzy neural networks [28].

Based on previous studies, we summarized the formation mechanism of frost hazards and the reasons for their uneven distribution in space (Figure 1). This study selected important factors (e.g., weather and terrain) and analyzed the relationship between factors affecting the occurrence of frost disaster and the hazard of frost disaster using artificial neural networks (ANNs) and support vector machines (SVMs) based on the case of 11 March 2017. This study also compared the accuracy of frost-occurrence models and constructed a frost-hazard model. Finally, the reliability of the model was verified using the yield data. The purpose of this study is to provide a model that can be used to analyze the spatial distribution of frost hazards for tea farmers in Zhejiang Province. In addition, combined with weather forecast data or climate change models, it can also be used to predict frost events in tea-planting areas.

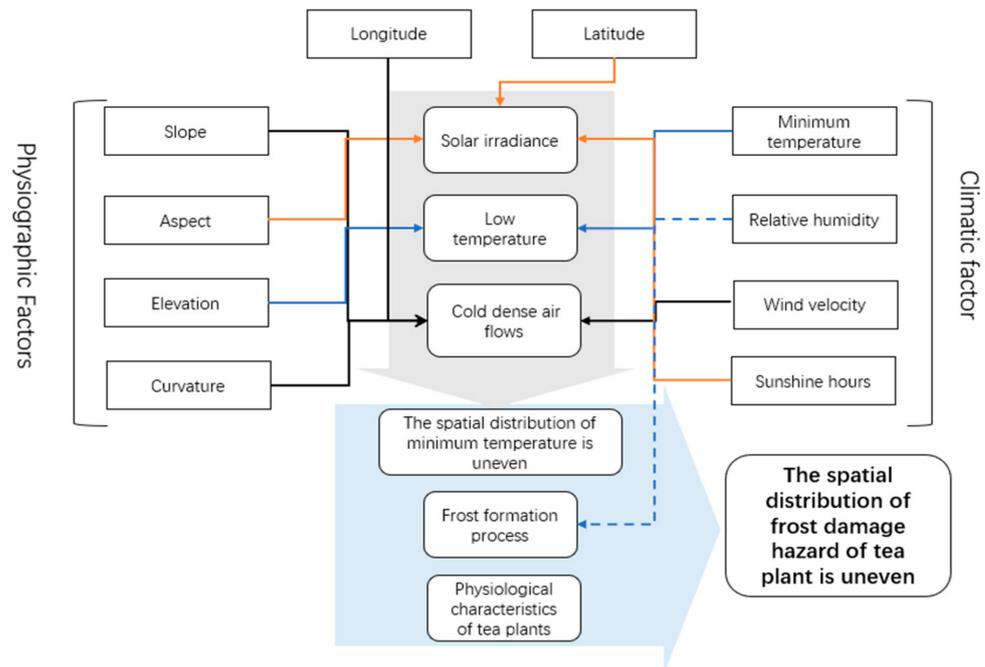


Figure 1. The formation of tea frost and the causes of its uneven spatial distribution.

2. Materials and Methods

2.1. Study Area

The study area (Zhejiang Province) is the main tea-planting area in China (Figure 2), with the country's highest tea export. Mean annual sunshine hours measure between 1600 and 2000 h, the frost-free period is >200 d, and the mean annual precipitation is between 1100 and 2000 mm. This "less-sunlight, warm, and humid" environment is highly suitable for tea-tree growth, and spring tea, with good quality and high economic benefit, is the main tree species planted by tea farmers [29]. In the past 20 years, Zhejiang Province has vigorously developed its famous, high-quality tea, and planted large areas of spring tea species. However, because mountainous or hilly terrain accounts for more than 70% of the total area in Zhejiang Province, and the transition zone in the middle and low latitudes is often affected by the monsoon; large-area tea trees often suffer from frost disasters in spring. Frost not only delays the growth of tea buds, reducing their price and quality, but also causes the death of tea buds, creating serious losses for farmers.

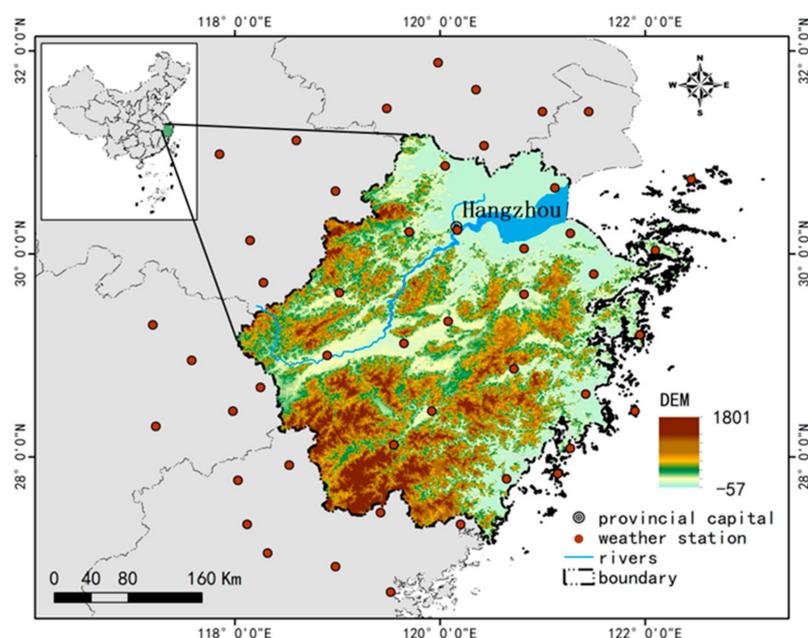


Figure 2. Study area. Note: The DEM in the legend is Digital Elevation Model, the unit is “meter”.

2.2. Data

The meteorological data in this study were obtained from the China Meteorological Data Network [30], which acquired daily meteorological data sets from 2000 to 2020, including the air minimum temperature, relative humidity, sunshine hours, wind speed, and other data recorded by 47 meteorological stations in Zhejiang Province and its surrounding areas. The Australian National University Spline (ANUSPLIN) package [31,32] and inverse distance weighted (IDW) interpolation were used to interpolate 47 meteorological stations. Digital elevation model (DEM) data were obtained from the Geospatial Data Cloud [33] using ArcGIS 10.4 (Environmental Systems Research Institute, The United States of America) spatial analysis tools to establish the slope, aspect, and curvature models. The county-level data of spring tea yield and planting area in Zhejiang Province from 1995 to 2019 (Huzhou, Quzhou, and Jinhua include only data from 2001 to 2019) were collected from Year Book China [34] to evaluate the accuracy of the model. Finally, combined with China’s land data in 2015, this study also generated the tea-planting area by visual interpretation based on the Landsat 8 remote-sensing image.

On 3 March 2017, the temperature dropped significantly in Zhejiang Province, and a tea frost event occurred. The Zhejiang meteorological station monitored this tea frost. According to the damage rate of buds and leaves, the frost grades of tea trees were classified as follows: the damage rate of buds and leaves $\leq 20\%$ was mild frost damage; $\sim 20\text{--}50\%$ was moderate frost damage; $\sim 50\text{--}80\%$ was severe frost damage; and $>80\%$ was serious frost damage. No damage and four frost grades were assigned as 0, 1, 2, 3, and 4, and selected samples in each frost grade were 32, 50, 82, 66, and 90, respectively; a total of 320 disaster points were used as model samples (Figure 3).

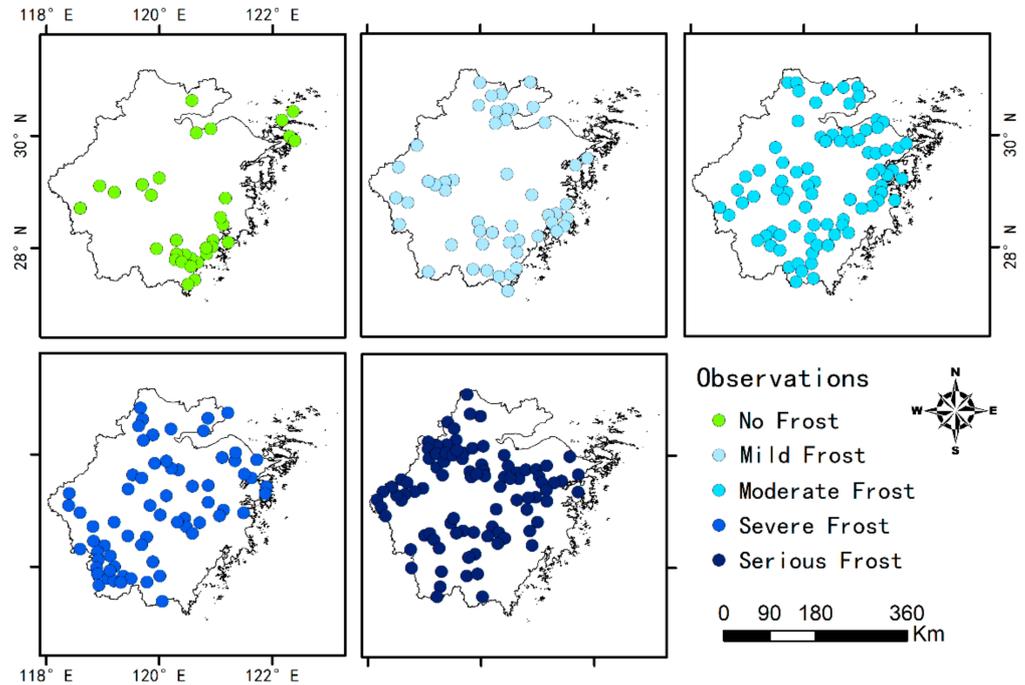


Figure 3. Collected records of occurrence points of spring frost damage on tea tree.

2.3. Methods

The study flowchart is shown in Figure 4. It includes three main parts: first, variables were selected to remove factors with high correlation and collinearity; second, the frost disaster point related to the factors, and the appropriate training and test samples were selected; and finally, the prediction model of spring frost hazard of tea plants in Zhejiang Province was established by SVM and ANN. The optimal model was selected by comparing the accuracy of the models, and the optimal model was used to predict the occurrence of a tea-tree frost event. The spatial distribution of frost hazard and its relationship with meteorological output determine the practical application of the model.

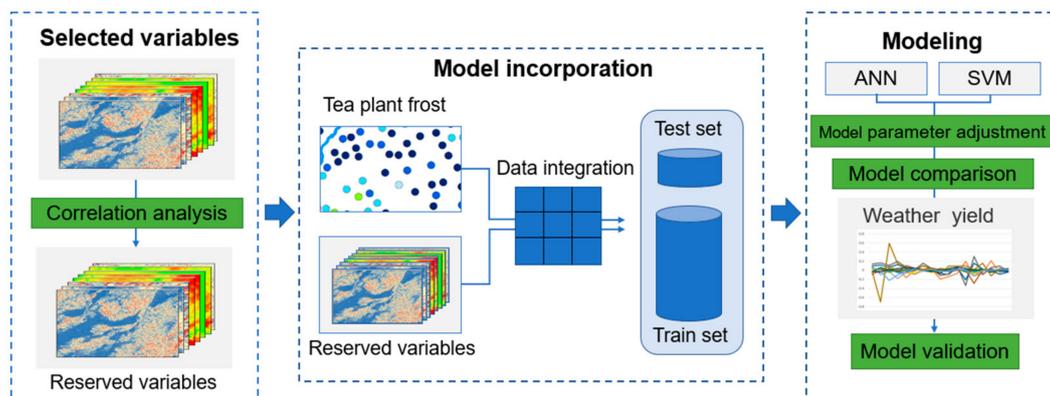


Figure 4. The flowchart of the study.

2.3.1. Artificial Neural Network

An artificial neural network (ANN) is a complex network structure formed by a large number of processing units (neurons) connected to each other. It is an information-processing system based on imitating the structure and function of brain neural networks

and simulating the activity of neurons using a mathematical model [35]. One of the main advantages of the ANN algorithm is its cost-effectiveness in real-time analysis, speed, and efficiency. It can also minimize errors and improve accuracy [36]. Therefore, it has become an important method for frost research. The formation and distribution of frost can be analyzed by predicting the minimum and dew-point temperatures. In the research of Chevalier et al., prediction of frost by ANN has been successfully applied to reality [37–39].

The back propagation (BP) algorithm of ANN is used mainly in research and includes two processes: forward propagation of the signal and backward propagation of error. In other words, the error output is calculated in the direction from input to output, while the weight and threshold are adjusted in the direction from output to input. In forward propagation, the input signal acts on the output node through the hidden layer, and the output signal is generated through a nonlinear transformation. If the actual output is not consistent with the expected output, it will turn into an error back-propagation process.

The number of hidden layer nodes influences the forecasting accuracy of the ANN. Having few nodes causes the network to learn less efficiently, requiring increased training times and affecting training accuracy; numerous nodes increase the training time, making the network easily overfit. The general formula for determining the number of hidden layer nodes l is:

$$l < \sqrt{(m + n) + a} \quad (1)$$

where m is the number of output layer nodes, n is the number of input layer nodes, and a is a constant between 0 and 10. The cut-and-try method was used to determine the optimal number of nodes so as to improve the performance of ANN.

2.3.2. SVM

A support vector machine (SVM) was first proposed by Vapnik; the main idea of the SVM is to establish a classification hyperplane as a decision surface to maximize the isolated edge between positive and negative examples [40]. An SVM is an approximate realization of structural risk minimization. The foundation and principle of the SVM is based on the fact that the error rate of machine learning on test data (i.e., generalization error rate) is bounded by the sum of the training error rate and a term depending on the Vapnik-Chervonenkis dimension. In the separable pattern case, the value of SVM for the first term is zero, and the second term is minimized. The unique attribute of the SVM in pattern classification is that it can provide better generalization performance.

The SVM algorithm was originally designed for binary classification problems. When dealing with multiclass classification problems, a classifier must be constructed directly or indirectly. In this research, we use the C-Support Vector Classifier (C-SVC) model, which is commonly used in SVM. In the multi-classification problem, the C-SVC model uses the “one-to-one” method, which designs an SVM between any two classes of samples, so $k(k - 1)/2$ SVM is needed for k classes of samples. When an unknown sample is classified, the category with the highest number of votes is the category of the unknown sample. The radial basis function kernel (RBF) with better performance is selected as the kernel function, which can map samples to higher spatial dimensions; it also requires fewer parameters, which reduces the difficulty of calculation [41]. The toolkit used in the study was LIBSVM, developed by Chih-Jen Lin et al. The latest version of LIBSVM is 3.25, which can be downloaded from <https://www.csie.ntu.edu.tw/~cjlin/libsvm/> (accessed on 4 June 2020) [42].

When using SVM to classify spring frost disaster events affecting tea plants, we need to adjust the penalty parameter C and kernel function parameter g to achieve good accuracy. Cross-validation (CV) is a statistical method often used to verify the performance of classifiers. The K-fold cross-validation (K-CV) in CV was selected and the original data were evenly divided into K ($K \geq 2$) groups. Each subset of data was used as a validation set, and the remaining $k-1$ subsets were used as the training sets. Finally, the mean value of the K model rows was selected as the output of the parameters. The advantage of this method is that it avoids overfitting or underfitting [43].

2.3.3. Methodologies for Model Evaluation

The performance of the two models has been evaluated under different conditions.

1. The Kappa coefficient can measure the accuracy of the multi-class classification problem when it is used in the consistency test, and its calculation method is based on a confusion matrix.

$$k = \frac{p_o - p_e}{1 - p_e} \quad (2)$$

where p_o represents the overall classification accuracy. Formula (3) is the calculation method of p_e . Assume that the number of real samples of each class is $a_1, a_2, a_3 \dots a_C$ (C is the number of classification categories. In our research, C is equal to 5) respectively, and the predicted number of samples of each class is $b_1, b_2, b_3 \dots b_C$ respectively, and the total sample size of the input model is n , then there are:

$$p_e = \frac{\sum_{i=1}^C a_i b_i}{n^2} \quad (3)$$

According to the previous experience, K usually falls between 0–1, which can be divided into five groups to represent the consistency of different levels, and generally when it falls between 0.61 and 0.80, it is considered to have a high degree of consistency [44].

2. Accuracy. This is the ratio of the number of correct samples to the total number of samples.
3. Average accuracy. This is the average accuracy of each sample. For imbalanced data, for n classes, the accuracy of each class is calculated respectively, and then the average value is calculated.

2.3.4. Meteorological Yield

In order to study the effect of spring frost disasters on tea yield, long-term trends in yield caused by human factors were first eliminated in this study (production level, policy, social economy, etc.). Yield (Y) time series can be decomposed into trend yield (Y_t), meteorological yield (Y_m) and random yield (ε), random production, (also called random noise), which is a random error term that can be ignored.

$$Y = Y_t + Y_m + \varepsilon \quad (4)$$

In this study, the quadratic exponential smoothing method was used to calculate the trend yield, which has been proved to be more universal in this study [45].

2.3.5. Selection of Variables

Many factors affect frost damage in tea trees, and they can be divided into two aspects: physiological and meteorological. The physiological aspect includes tea varieties, growth stages, plant age, branch and leaf maturity, and picking level. The meteorological aspect depends on the intensity and duration of low temperature, as well as wind direction, wind speed, air, and soil humidity. At the same time, longitude and latitude can also affect the spatial distribution of temperature through meteorological factors [46,47]. Moreover, altitude can affect the vertical distribution of air temperature, and local topography can further affect the flow path and convergence of cold air, leading to an uneven spatial distribution of minimum temperature [19,20,48].

Combined with previous studies [25,27,49], this study selected 10 variables, including longitude and latitude, meteorological factors (e.g., relative humidity, wind velocity, sunshine hours, and minimum temperature), and terrain factors (e.g., elevation, aspect, slope, and curvature). Then, two types of tests were conducted to verify the correlation and collinearity between independent variables. Pearson correlation coefficients of any two of the 10 variables were calculated to eliminate variables with high correlation. To avoid high multicollinearity among the selected variables, which may lead to a large deviation in the

classification accuracy of the model, and to select variables with better independence and higher explanatory ability, the coefficient of variance inflation factor (VIF) was used to test the linear correlation between factors [50].

The correlation analysis shows that the Pearson correlation coefficients of the longitude, sunshine, and wind speed were 0.725 and 0.667, respectively, indicating a strong correlation between them. Since the correlation of sunshine and wind speed with other variables was less than ± 0.3 , the longitudinal variables were excluded from the study. In previous studies, VIF values > 4 were regarded as evidence of multicollinearity [25]. In the process of calculation, it was found that the largest VIF value of the remaining variables was 2.134 of the aspect variable (Table 1); therefore, nine variables were retained, excepting longitude.

Table 1. Selected variables in the study.

Variable	Min	Max	VIF
Latitude	118.156	30.737	1.684
Slope	0	50.306	1.047
Aspect	−1	359.963	2.134
Elevation	−57	1801	1.038
Curvature	−0.210	0.231	1.1
Minimum temperature	−1.856	7.647	1.187
Relative humidity	31.323	69.710	1.881
Sunshine hours	4.420	9.692	2.082
Wind velocity	0.843	4.232	1.239

3. Results

3.1. Model Parameter Adjustment

We apply ANN and SVM to a group of variables. The data were divided into training samples (75%) and test samples (25%). Test samples are used for unbiased evaluation of the final fitting of the model on the sample training data. As a group of data not involved in the construction of the model, the generalization of the model can be tested, and the accuracy of the model can be calculated by testing the actual and predicted values of the samples.

The construction of SVM model needs to adjust parameters c and g . The parameters c and g of the model are 194.012 and 0.144, respectively, after the samples are input into the model for the first time, and the accuracy is 80.417%. Then, taking the training set as the original data, the best parameters c and g are obtained by K-CV method. In the process of parameter selection, there may be multiple sets of c and g corresponding to the highest accuracy of verification classification. In this part, we choose a group of c and g (our results are 64 and 0.5) which can reach the lowest parameter c under the condition of the highest classification accuracy as the best parameters (Figure 5). This avoids the problem of over fitting caused by too high penalty parameters, reduces the generalization ability of the classifier, and improves the accuracy to 83.75%.

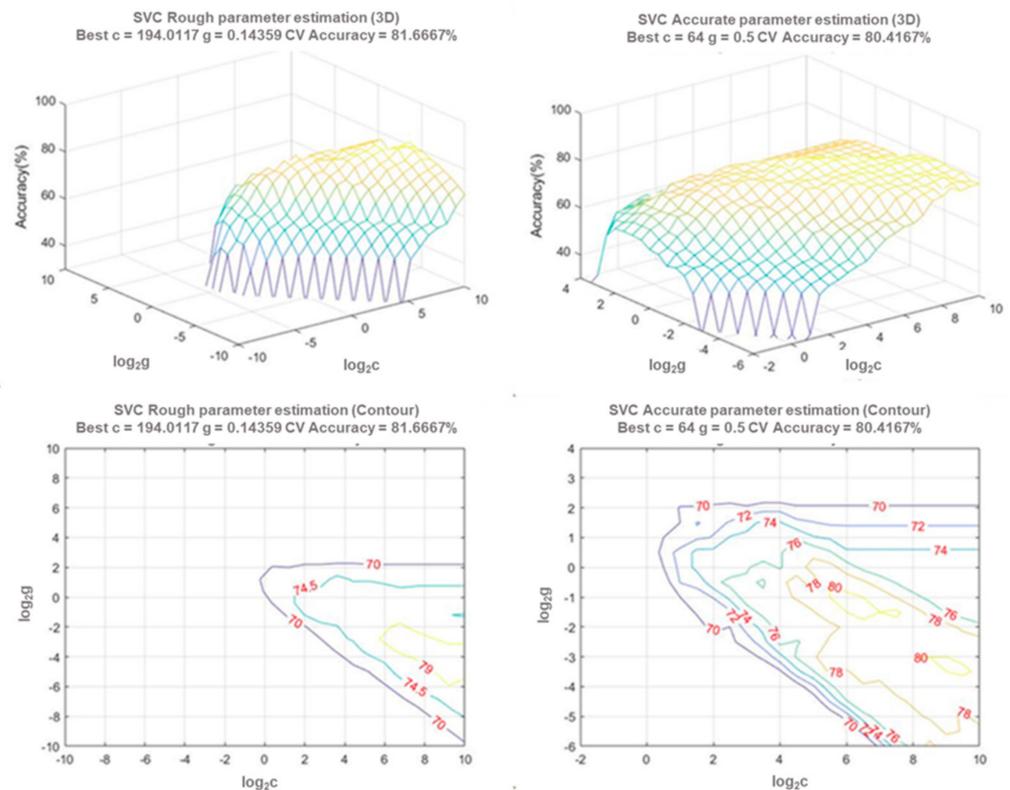


Figure 5. Taking model as an example, the 3D view and contour map of the precision process of model parameters c and g are shown, respectively.

The number of hidden layers and hidden layer nodes of the neural network have an influence on the classification results. As for the number of hidden layers, Robert [51] proved theoretically that any continuous function in a closed interval can be approximated by a BP network with a hidden layer, so a three-layer BP network can complete any mapping from m dimension to m dimension. Increasing the number of layers can improve the learning accuracy, but on the other hand, it also makes the network structure complex and increases the training time [52]. Therefore, we only consider the single hidden layer neural network model.

Figure 6 shows the influence of the different numbers of hidden layer nodes on the error rate of model classification; the error rate is lowest when l is equal to 8. In addition, the classification accuracy is 75%.

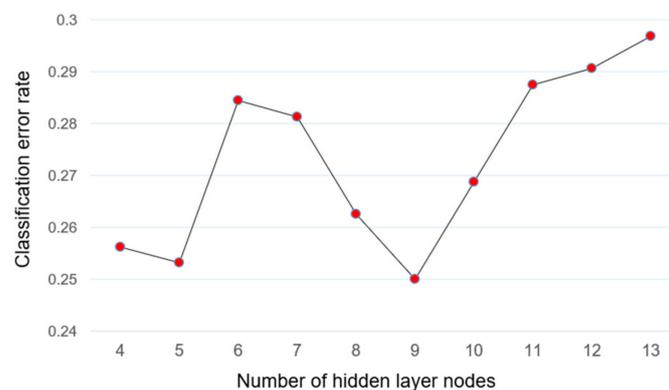


Figure 6. The error rate changes.

3.2. Classification Results

The accuracy rate of the results obtained by the SVM model was 83.75%, and the sensitivities of different categories of models were quite different. The correct classification rate of Category 0 was only 37.5%, while that of Category 3 was as high as 100%. In addition, there was slight confusion in Categories 2 and 4, with 25% of the samples in Category 2 classified into Category 4, and 8.3% of the samples in Category 4 classified into Category 2 (Figure 7).

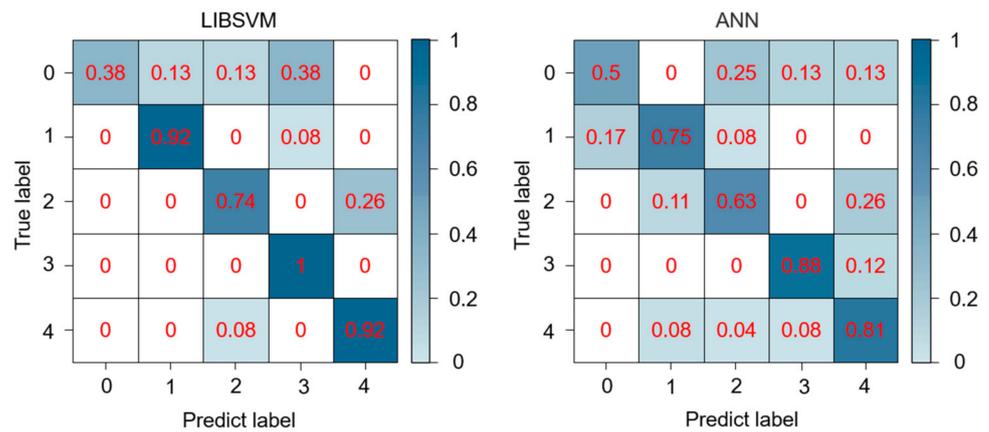


Figure 7. Confusion matrix of classification results. Note: Because the values in the graph are rounded to two decimal places, some values are larger or smaller than the original values, resulting in the sum of each row not being equal to one.

The overall accuracy of the ANN model is lower than that of SVM, and it is better than SVM in 0 category classification, but it is not ideal. In addition, the classification results of Categories 1–4 are worse than those of SVM. The same as SVM is that the recognition accuracy of Categories 3 and 4 is high, which is related to the large number of training samples.

The accuracy, average accuracy and Kappa coefficient were calculated for each of the two models. The results show that the evaluation result of the SVM model is better than that of the ANN model (Table 2).

Table 2. Model classification performance.

	Accuracy	Average Accuracy	Kappa Coefficient
SVM	0.8375	0.7929	0.791
ANN	0.75	0.7129	0.6737

3.3. Actual Prediction of the Models

In this study, the SVM model was verified to be more suitable for predicting the hazard of spring frost to tea trees in Zhejiang Province. In order to verify the practical application of the model, the spatial distribution of the frost-hazard degree of tea trees in Zhejiang Province in 2016 was predicted:

The late-spring cold event on 11 March 2016, caused frost damage to more than half of the tea plantations in Zhejiang Province, covering an area of more than 100,000 hectares. Approximately 3700 tons of early tea were damaged, with an estimated economic loss of 1.8 billion yuan. Lishui, Hangzhou, Shaoxing, Huzhou, and other places suffered the most serious damage. In this study, the economic crop forests based on land use types in Zhejiang Province were selected as the sample points, and nine research elements corresponding to the sample points were extracted as variables and input into the model to obtain the classification results of the model.

From overall classification results, the disasters in the east of Zhejiang Province and the Yangtze River estuary are shown to be at a lower level, due to the fact that a body of water can adjust and compensate for temperature when encountering strong cold air, raising the extreme minimum temperature, and correspondingly reducing the harm from cold spells in late spring. Due to the undulating terrain, the air temperature in the southwest and northwest decreases vertically; the influence of frost events is closely related to altitude, slope aspect, and other topographic factors. In the central basin, due to its high density, cold air deposits in low-lying areas, so it more readily causes serious frost events in these low areas than in plain areas. The accuracy of the classification results can be proved using empirical theory. However, the correlation between the results and the meteorological yield was further analyzed to determine the accuracy of the model.

The gray areas in Figure 8 are the main spring tea planting counties in Zhejiang Province. The samples of each county were counted, the mean value of frost grade (M) of the area is calculated according to the Equation 5, and the calculation results are shown in Table 3.

$$M = \frac{\sum_{i=1}^5 i \times x_i}{n} \quad (5)$$

where i is the assignment of frost damage level of tea tree corresponding to the sample, x_i is the number of sample points corresponding to class i in the region, and n is the number of sample points.

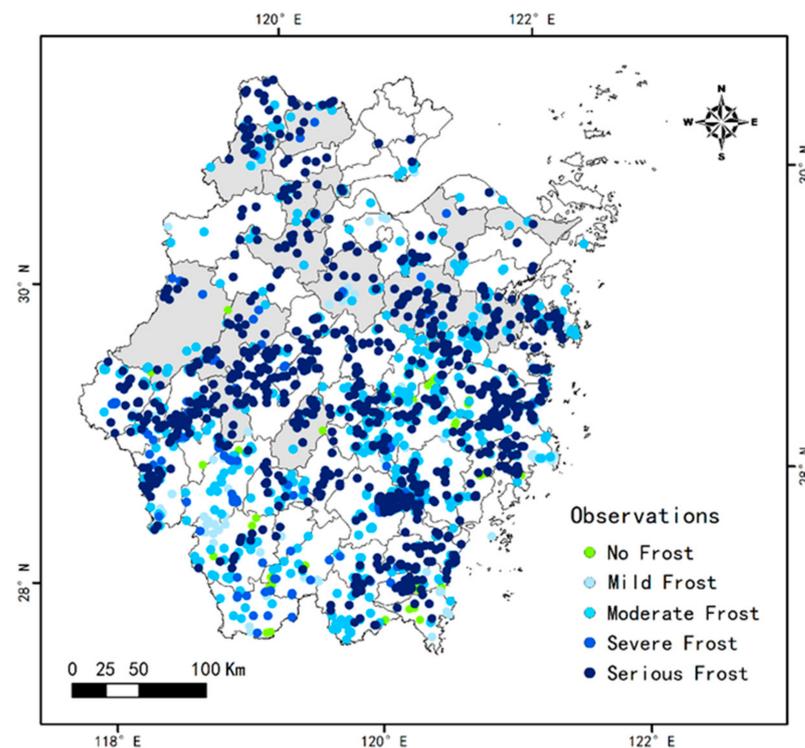


Figure 8. Model prediction results of the study. Note: For convenience of observation, raster data are converted to a vector point display.

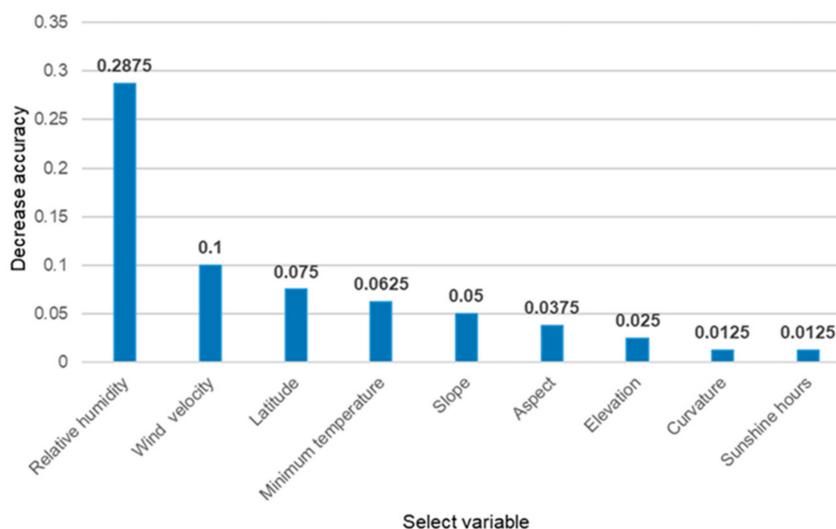
By calculating the correlation between M and meteorological yield, Pearson correlation coefficient was found to be -0.79 ($p < 0.01$) (two tailed), that is, there is a good negative correlation between the regional frost level average output of the model and the meteorological yield of tea, which indicates that the model also has a high fitting effect in actual production.

Table 3. Mean value of meteorological yield and regional frost hazard grade of spring tea in main planting areas.

Area	Meteorological Yield	M	Area	Meteorological Yield	M
Hang zhou	−0.017	3.000	Ji an	0.005	3.071
Sheng zhou	−0.055	3.560	Xin chang	−0.086	3.241
Hu zhou	0.029	3.067	Ning bo	0.045	2.667
Wu yi	−0.040	3.385	Yu yao	0.045	2.429
Yu hang	0.062	2.760	Ning hai	−0.018	3.027
Zhu ji	0.149	2.696	Long you	−0.049	3.360
Fu yang	−0.124	3.529	Jian de	−0.012	3.094
Chun an	−0.012	2.777			

3.4. Factor Importance Analysis

Since several input variables may affect the hazard of tea-tree frost damage in different ways, it is necessary to study the importance and mechanism of variables in order to provide better guidance for the prediction and control of frost damage to tea trees. In this study, the importance of nine variables was evaluated using the sensitivity analyses of the SVM model [53]. That is, the degree to which the prediction accuracy of the predictor is reduced by removing the factors one-by-one, and the prediction accuracy of the model before and after elimination, were analyzed. The importance of the factors increases with their differences [54]. The order of importance of condition factors is shown in Figure 9.

**Figure 9.** Factor importance ranking.

Relative humidity is the most important factor affecting the risk of spring frost damage to tea trees in the study area, and the average precision is reduced by 0.2875. The wind speed is less important, and the average precision is reduced by 0.1. The two least important adjustment factors are curvature (0.0125) and sunshine hours (0.0125).

4. Discussion

In this study, we found that air humidity is the most important factor affecting the accuracy of prediction, followed by wind speed, latitude, minimum temperature, terrain factors, and solar radiation. The spring in Zhejiang Province is the alternating season between the East Asian monsoon in the winter and the summer monsoon, with frequent north–south airflow exchange, low air pressure, and cold and rainy fronts, so air humidity is maintained over a large duration. Air humidity affects the supercooling temperature

of the plants [55]. Frost occurs when the temperature is lower than the supercooling temperature. Although there is little information on the supercooling temperature in the study of tea plants, it is undeniable that it is meaningful to study the influence of relative humidity on the occurrence of frost. Damage by frost can be reduced by reducing the air humidity or spraying humidification. In Zhejiang Province, which is dominated by radiation frost, frost usually occurs in the early morning or at night, and when the air forms a stable radiation inversion to form frost; however, if the wind speed is high, this can disturb the inversion layer structure and reduce the degree of damage incurred by frost. Based on this conclusion [11], we believe that wind turbines are an effective means of preventing and controlling spring frost damage to tea trees in Zhejiang Province.

Latitude and longitude affect the degree of frost damage on a large scale. Longitude affects the location of land and sea, thus affecting the continental characteristics of the region. However, the study area is located on the southeast coast of China, and the entire region is significantly affected by the Pacific monsoon climate. We also tried to input longitude as a variable into the model and found that the classification accuracy of the model was reduced by 5–8%. However, in regions with significant continental characteristics, longitude is an important variable [56]. Latitude is the third most important factor in our study, affecting the zonal distribution of temperature, thus affecting the risk of frost.

The minimum temperature is generally considered to be the most important factor affecting frost occurrence [57,58]. In previous studies, scientists used the threshold and duration of the minimum temperature to assess the severity of frost [16,57]. However, in our study, the minimum temperature was not the most important factor affecting the accuracy of the model. The results of Kotikot et al. [18] show that there is a negative correlation between the surface temperature and the occurrence of frost (0 indicates frost-free, 1 indicates frost), but its weight is lower than that of most terrain factors. This shows that in the local tea-planting area, the air temperature cannot completely reflect the occurrence of frost due to the influence of microclimate and terrain on the accumulation of cold air. The difference between our study and other studies is that frost does not occur in one of the five classification labels. Even if the occurrence of frost may be greatly affected by low temperatures, the influence of minimum temperature on different degrees of severity may be low, and this requires further study and discussion.

In selection of the model, we choose ANN and SVM, which have the advantages of simple operation, low learning cost, and stability. However, the results of the ANN are not ideal. Since ANN classification may be affected by the workload of the provided samples, the variables transmit calculation results to the output layer nodes through the hidden layer, and finally adjust the weight to reduce the sum of the prediction square errors of the dependent variables; therefore, the input and output of the training samples affect the classification performance of the neural network. In our research, SVM is a better choice, but we did not try a random forest model, Bayesian, or more a complex artificial neural network, which should be addressed in future research.

5. Conclusions

In this study, based on ANN and SVM model, the spring frost hazard for tea tree in Zhejiang Province was classified and modeled. The machine learning classifier was used to combine the spring frost on tea tree with terrain and meteorological factors, and the selected nine factors were input into the model as variables. By adjusting the parameters, the best classification model with the highest accuracy was obtained. By calculating the accuracy and consistency of the model, it was concluded that the SVM model was more suitable. Then, SVM model was applied to predict the spatial distribution of the frost hazard to tea tree in 2016. The results showed that the frost damage was more serious in the middle and the north due to the influence of topography and latitude, respectively, and the degree of frost damage was lower in the coastal and river areas because of the buffering effect of water at low temperatures. Compared with previous studies on frost

prediction, this study predicts different hazard levels of frost disasters of tea trees, so that the producers can not only distinguish the occurrence and nonoccurrence of frost through the model, but also respond according to different levels. In this region, the model has high classification accuracy and reliability, which is conducive to improving the effectiveness of frost prevention and saving resources.

Finally, we analyzed the importance of different factors and identified the most important factors affecting the spring frost damage to tea trees in the study area. Relative humidity and wind speed were found to be important factors affecting the classification of the model in our study, indicating that spring frost damage to tea trees in the study area is related mainly to the relative humidity and wind speed, and that use of fans and water spray may effectively reduce the damage caused by frost to tea trees.

Author Contributions: Conceptualization, J.X. and G.R.; methodology, J.X. and X.L.; software, J.X.; validation, G.R. and S.G.; data curation, J.X. and K.L.; writing—original draft preparation, J.X.; writing—review and editing, S.G. and D.R.; visualization, S.G.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Key Research and Development Program of China (2019YFD1002201); the National Natural Science Foundation of China (41877520); the Science and Technology Development Planning of Jilin Province (20190303018SF); the Key Research and Projects Development Planning of Jilin Province (20200403065SF); the Science and Technology Planning of Changchun (19SS007).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We appreciate the editors and the reviewers for their constructive suggestions and insightful comments, which helped greatly to improve this manuscript.

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

References

1. FAO. International Tea Day. *Tea: Resilient, Sustainable and Healthy from Field to Cup*. Available online: <http://www.fao.org/international-tea-day/en/> (accessed on 11 June 2021).
2. Carr, M.K.V. The Climatic Requirements of the Tea Plant: A Review. *Exp. Agric.* **1972**, *8*, 1–14. [CrossRef]
3. Chen, H.; Liu, C.; Liu, C.; Hu, C.; Hsiao, M.; Chiou, M.; Su, Y.; Tsai, H. A Growth Model to Estimate Shoot Weights and Leaf Numbers in Tea. *Biometry Modeling Stat.* **2019**, *111*, 2255–2262. [CrossRef]
4. Lu, Y.; Hu, Y.; Snyder, R.L.; Kent, E.R. Tea leaf's microstructure and ultrastructure response to low temperature in indicating critical damage temperature. *Inf. Process. Agric.* **2019**, *6*, 247–254. [CrossRef]
5. Wang, F.; Zhang, Y.Q. The influence of late spring coldness on famous tea production and its prevention. *Agric. Equip. Technol.* **2004**, *26*. [CrossRef]
6. Lou, W.; Sun, K.; Sun, S.; Ma, F.; Wang, D. Changes in pick beginning date and frost damage risk of tea tree in Longjing tea-producing area. *Theor. Appl. Clim.* **2013**, *114*, 115–123. [CrossRef]
7. Ma, S.Q.; Li, F.; Wang, Q. *Cold Wave and Frost*; Meteorological Press: Beijing, China, 2009; pp. 67–82.
8. Lu, Y.Z. Response Characteristics of Radiation Frost in Tea Fields and Energy Quantitative for Frost Protection. Ph.D. Thesis, Jiangsu University, Zhenjiang, China, 2020; pp. 14–21.
9. Lu, Y.; Hu, Y.; Li, P.; Paw U, K.T.; Snyder, R.L. Prediction of Radiation Frost Using Support Vector Machines Based on Micrometeorological Data. *Appl. Sci.* **2019**, *10*, 283. [CrossRef]
10. Liu, J.; Chen, N. Investigation and prevention of frost damage in tea garden. *Newsl. Seric. Tea* **2003**, *3*, 12–13. [CrossRef]
11. Snyder, R.L.; Paw, U.K.T.; Thompson, J.F. *Frost Protection: Fundamentals, Practice, and Economics*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2005; Volume 1, pp. 1–10.
12. Sgubin, G.; Swingedouw, D.; Dayon, G.; de Cortázar-Atauri, I.G.; Ollat, N.; Pagé, C.; van Leeuwen, C. The risk of tardive frost damage in French vineyards in a changing climate. *Agric. For. Meteorol.* **2018**, *250–251*, 226–242. [CrossRef]
13. Xiao, L.; Liu, L.; Asseng, S.; Xia, Y.M.; Tang, L.; Liu, B.; Cao, W.; Zhu, Y. Estimating spring frost and its impact on yield across winter wheat in China. *Agric. For. Meteorol.* **2018**, *260*, 154–164. [CrossRef]
14. Wang, S.; Chen, J.; Rao, Y.H.; Liu, L.C.; Wang, W.Q.; Dong, Q. Response of winter wheat to spring frost from a remote sensing perspective: Damage estimation and influential factors. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 221–235. [CrossRef]

15. Zhao, L.; Li, Q.; Zhang, Y.; Wang, H.; Du, X. Normalized NDVI valley area index (NNVAI)-based framework for quantitative and timely monitoring of winter wheat frost damage on the Huang-Huai-Hai Plain, China. *Agric. Ecosyst. Environ.* **2020**, *292*, 106793. [[CrossRef](#)]
16. Chatrabgoun, O.; Karimi, R.; Daneshkhah, A.; Abolfathi, S.; Nouri, H.; Esmailbeigi, M. Copula-based probabilistic assessment of intensity and duration of cold episodes: A case study of Malayer vineyard region. *Agric. For. Meteorol.* **2020**, *295*, 108150. [[CrossRef](#)]
17. Kotikot, S.M.; Flores, A.; Griffin, R.E.; Sedah, A.; Nyaga, J.; Mugo, R.; Limaye, A.; Irwin, D.E. Mapping threats to agriculture in East Africa: Performance of MODIS derived LST for frost identification in Kenya's tea plantations. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 131–139. [[CrossRef](#)]
18. Wang, P.; Ma, Y.; Tang, J.; Wu, D.; Chen, H.; Jin, Z.; Huo, Z. Spring Frost Damage to Tea Plants Can Be Identified with Daily Minimum Air Temperatures Estimated by MODIS Land Surface Temperature Products. *Remote Sens.* **2021**, *13*, 1177. [[CrossRef](#)]
19. Alan, E.L. Effects of slope and aspect variations on satellite surface temperature retrievals and mesoscale analysis in mountainous terrain. *J. Appl. Meteorol. Climatol.* **1992**, *31*, 255–264. [[CrossRef](#)]
20. Gerlitz, L. Using fuzzified regression trees for statistical downscaling and regionalization of near surface temperatures in complex terrain. *Theor. Appl. Clim.* **2014**, *122*, 337–352. [[CrossRef](#)]
21. Kerdiles, H.; Grondona, M.; Rodriguez, R.; Seguin, B. Frost mapping using NOAA AVHRR data in the Pampean region, Argentina. *Agric. For. Meteorol.* **1996**, *79*, 157–182. [[CrossRef](#)]
22. Lindkvist, L.; Gustavsson, T.; Bogren, J. A frost assessment method for mountainous areas. *Agric. For. Meteorol.* **2000**, *102*, 51–67. [[CrossRef](#)]
23. Pouteau, R.; Rambal, S.; Ratte, J.-P.; Gogé, F.; Joffre, R.; Winkel, T. Downscaling MODIS-derived maps using GIS and boosted regression trees: The case of frost occurrence over the arid Andean highlands of Bolivia. *Remote Sens. Environ.* **2011**, *115*, 117–129. [[CrossRef](#)]
24. Kotikot, S.M.; Onywere, S.M. Application of GIS and remote sensing techniques in frost risk mapping for mitigating agricultural losses in the Aberdare ecosystem, Kenya. *Geocarto Int.* **2014**, *30*, 104–121. [[CrossRef](#)]
25. Kotikot, S.M.; Flores, A.; Griffin, R.E.; Nyaga, J.; Case, J.L.; Mugo, R.; Sedah, A.; Adams, E.; Limaye, A.; Irwin, D.E. Statistical characterization of frost zones: Case of tea freeze damage in the Kenyan highlands. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *84*, 101971. [[CrossRef](#)]
26. Gobbett, D.L.; Nidumolu, U.; Crimp, S. Modelling frost generates insights for managing risk of minimum temperature extremes. *Weather Clim. Extrem.* **2020**, *27*, 100176. [[CrossRef](#)]
27. Lee, H.; Chun, J.A.; Han, H.-H.; Kim, S. Prediction of Frost Occurrences Using Statistical Modeling Approaches. *Adv. Meteorol.* **2016**, *2016*, 1–9. [[CrossRef](#)]
28. Yue, Y.; Zhou, Y.; Wang, J.; Ye, X. Assessing Wheat Frost Risk with the Support of GIS: An Approach Coupling a Growing Season Meteorological Index and a Hybrid Fuzzy Neural Network Model. *Sustainability* **2016**, *8*, 1308. [[CrossRef](#)]
29. Meng, Z. Effect of low temperature and frost on April 1 on tea product in Zhejiang. *J. Zhejiang Agric. Sci.* **2019**, *60*, 1397–1400. [[CrossRef](#)]
30. China meteorological data network. Available online: <http://data.cma.cn/> (accessed on 27 June 2021).
31. Liu, Z.H.; Tim, R.M.; Tom, G.V.; Yang, K.Q.; Li, R.; Mu, X.M. Interpolation for time series of meteorological variables using ANUSPLIN. *J. Northwest A&F Univ.* **2008**, *36*, 227–234. [[CrossRef](#)]
32. Hutchinson, M.F. ANUSPLIN Version 4.3 User Guide, The Australian National University, Centre for Resource and Environmental Studies. Canberra. Available online: <https://fennerschool.anu.edu.au/research/products/anusplin> (accessed on 22 May 2020).
33. Geospatial Data Cloud. Available online: <http://www.gscloud.cn/> (accessed on 27 June 2021).
34. Year Book China. Available online: <https://www.yearbookchina.com/> (accessed on 27 June 2021).
35. Wu, C.Y. The Research and Application on Neural Network. Master's Thesis, Northeast Agricultural University, Harbin, China, 2007.
36. Robinson, C.; Mort, N. A neural network system for the protection of citrus crops from frost damage. *Comput. Electron. Agric.* **1997**, *16*, 177–187. [[CrossRef](#)]
37. Shank, D.B.; Hoogenboom, G.; McClendon, R.W. Dewpoint Temperature Prediction Using Artificial Neural Networks. *J. Appl. Meteorol. Clim.* **2008**, *47*, 1757–1769. [[CrossRef](#)]
38. Chevalier, R.F.; Hoogenboom, G.; McClendon, W.R.; Paz, J.O. A web-based fuzzy expert system for frost warnings in horticultural crops. *Environ. Model. Softw.* **2012**, *35*, 84–91. [[CrossRef](#)]
39. Chevalier, R.F.; Hoogenboom, G.; McClendon, W.R.; Paz, J.A. Support vector regression with reduced training sets for air temperature prediction: A comparison with artificial neural networks. *Neural Comput. Appl.* **2011**, *20*, 151–159. [[CrossRef](#)]
40. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
41. Brereton, R.G.; Lloyd, G.R. Support Vector Machines for classification and regression. *Analyst* **2010**, *135*, 230–267. [[CrossRef](#)]
42. Hsu, C.W.; Chang, C.C.; Lin, C.J. A Practical Guide to Support Vector Classification, Department of Computer Science National Taiwan University, 2003. Taipei 106, Taiwan. Available online: <http://www.csie.ntu.edu.tw/~cjlin/libsvm> (accessed on 4 June 2020).
43. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27. Available online: <http://www.csie.ntu.edu.tw/~cjlin/libsvm> (accessed on 4 June 2020). [[CrossRef](#)]

44. Tang, W.; Hu, J.; Zhang, H.; Wu, P.; He, H. Kappa coefficient: A popular measure of rater agreement. *Shanghai Arch. Psychiatry* **2015**, *27*, 62–67. [[CrossRef](#)]
45. Li, X.Y.; Zhang, Y.; Zhao, Y.X.; Du, Z.Y.; Yang, S. Comparative study on main crop yield separation methods. *J. Appl. Meteorol. Sci.* **2020**, *31*, 74–82. [[CrossRef](#)]
46. Huang, S.B. Meteorology of the tea plant in China: A review. *Agric. For. Meteorol.* **1989**, *47*, 19–30. [[CrossRef](#)]
47. Layomi Jayasinghe, S.; Kumar, L.; Sandamali, J. Assessment of Potential Land Suitability for Tea (*Camellia sinensis* (L.) O. Kuntze) in Sri Lanka Using a GIS-Based Multi-Criteria Approach. *Agriculture* **2019**, *9*, 148. [[CrossRef](#)]
48. Alvares, C.A.; Sentelhas, P.C.; Stape, J.L. Modeling monthly meteorological and agronomic frost days, based on minimum air temperature, in Center-Southern Brazil. *Theor. Appl. Climatol.* **2017**, *134*, 177–191. [[CrossRef](#)]
49. Hengl, T.; Heuvelink, G.B.M.; Tadić, M.P.; Pebesma, E. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. *Theor. Appl. Climatol.* **2011**, *107*, 265–277. [[CrossRef](#)]
50. Fox, J.; Monette, G. Generalized Collinearity Diagnostics. *J. Am. Stat. Assoc.* **1992**, *87*, 178–183. [[CrossRef](#)]
51. Hecht-Nielsen, R. Theory of the backpropagation neural network. *Neural Netw.* **1988**, *1*, 445. [[CrossRef](#)]
52. Jiao, B.; Ye, M.X. Determination of hidden unit number in a BP neural network. *J. Shanghai Dianji Univ.* **2013**, *16*, 113–116. [[CrossRef](#)]
53. Du, G.L.; Zhang, Y.-S.; Iqbal, J.; Yang, Z.-H.; Yao, X. Landslide susceptibility mapping using an integrated model of information value method and logistic regression in the Bailongjiang watershed, Gansu Province, China. *J. Mt. Sci.* **2017**, *14*, 249–268. [[CrossRef](#)]
54. Sun, D.L.; Wen, H.J.; Wang, D.Z.; Xu, J.H. A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology* **2020**, *362*, 107201. [[CrossRef](#)]
55. Wang, J.; Li, H.; Ma, G.; Duan, X.; Zhang, X. Effects of air humidity on super cooling point of Fuji apple flowers. *Non Wood For. Res.* **2020**, *38*, 225–230. [[CrossRef](#)]
56. Écio Souza, D.; Lorenzon, S.A.; de Castro, N.L.M.; Marcatti, G.E.; dos Santos, O.P.; Júnior, J.C.D.D.; Cavalcante, R.B.L.; Fernandes-Filho, E.I.; Amaral, C.H. Forecasting frost risk in forest plantations by the combination of spatial data and machine learning algorithms. *Agric. For. Meteorol.* **2021**, *306*, 108450. [[CrossRef](#)]
57. Ghielmi, L.; Eccel, E. Descriptive models and artificial neural networks for spring frost prediction in an agricultural mountain area. *Comput. Electron. Agric.* **2006**, *54*, 101–114. [[CrossRef](#)]
58. Ding, L.; Noborio, K.; Shibuya, K. Frost Forecast using Machine Learning—From association to causality. *Procedia Comput. Sci.* **2019**, *159*, 1001–1010. [[CrossRef](#)]