



Article

Remote Sensing Monitoring of Drought in Southwest China Using Random Forest and eXtreme Gradient Boosting Methods

Xiehui Li ^{1,2,3,*}, Hejia Jia ^{1,4} and Lei Wang ¹¹ School of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu 610225, China² Yunnan R&D Institute of Natural Disaster, Chengdu University of Information Technology, Kunming 650034, China; jhj_1221@outlook.com (H.J.); lwang@cuit.edu.cn (L.W.)³ Key Open Laboratory of Arid Climate Change and Disaster Reduction, China Meteorological Administration/Key Laboratory of Arid Climatic Change and Reducing Disaster of Gansu Province, Lanzhou Institute of Arid Meteorology, China Meteorological Administration, Lanzhou 730020, China⁴ Xianning Meteorological Service, Xianning 437000, China

* Correspondence: lixiehui@cuit.edu.cn or lixiehui325328@163.com

Abstract: A drought results from the combined action of several factors. The continuous progress of remote sensing technology and the rapid development of artificial intelligence technology have enabled the use of multisource remote sensing data and data-driven machine learning (ML) methods to mine drought features from different perspectives. This method improves the generalization ability and accuracy of drought monitoring and prediction models. The present study focused on drought monitoring in southwest China, where drought disasters occur frequently and with a high intensity, especially in areas with limited meteorological station coverage. Several drought indices were calculated based on multisource satellite remote sensing data and weather station observation data. Remote sensing data from multiple sources were combined to build a reconstructed land surface temperature (LST) and drought monitoring method using the two different ML methods of random forest (RF) and eXtreme Gradient Boosting (XGBoost 1.5.1), respectively. A 5-fold cross-validation (CV) method was used for the model's hyperparameter optimization and accuracy evaluation. The performance of the model was also assessed and validated using several accuracy assessment indicators. The model monitored the results of the spatial and temporal distributions of the drought, drought grades, and influence scope of the drought. These results from the model were compared against historical drought situations and those based on the standardized precipitation evapotranspiration index (SPEI) and the meteorological drought composite index (MCI) values estimated using weather station observation data in southwest China. The results show that the average score of the 5-fold CV for the RF and XGBoost was 0.955 and 0.931, respectively. The root-mean-square error (RMSE) of the LST values reconstructed using the RF model on the training and test sets was 1.172 and 2.236, the mean absolute error (MAE) was 0.847 and 1.719, and the explained variance score (EVS) was 0.901 and 0.858, respectively. Furthermore, the correlation coefficients (CCs) were all greater than 0.9. The RMSE of the monitoring values using the XGBoost model on the training and test sets was 0.135 and 0.435, the MAE was 0.095 and 0.328, the EVS was 0.976 and 0.782, and the CC was 0.982 and 0.868, respectively. The consistency rate between the drought grades identified using SPEI1 (the SPEI values of the 1-month scale) based on the observed data from the 144 meteorological stations and the monitoring values from the XGBoost model was more than 85%. The overall consistency rate between the drought grades identified using the monitoring and MCI values was 67.88%. The aforementioned two different ML methods achieved a high comprehensive performance, accuracy, and applicability. The constructed model can improve the level of dynamic drought monitoring and prediction for regions with complex terrain and topography and formative factors of climate as well as where weather stations are sparsely distributed.

Keywords: drought index; drought monitoring; RF model; XGBoost model; multisource remote sensing information; southwest China



Citation: Li, X.; Jia, H.; Wang, L. Remote Sensing Monitoring of Drought in Southwest China Using Random Forest and eXtreme Gradient Boosting Methods. *Remote Sens.* **2023**, *15*, 4840. <https://doi.org/10.3390/rs15194840>

Academic Editor: Luca Brocca

Received: 27 July 2023

Revised: 6 September 2023

Accepted: 20 September 2023

Published: 6 October 2023



Copyright: © 2023 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/>).

1. Introduction

A continual global warming trend, characterized by an increase in both the frequency and intensity of extreme weather and climate events, has been witnessed in recent years. This warming trend has significantly impacted human survival and sustainable socio-economic development. Drought is one of the natural disasters that affects the most extensive amount of area and causes the greatest losses [1–3]. Drought has become increasingly frequent worldwide because of the dual effects of climate change and human activities, and it usually occurs sequentially and concurrently. According to the Annual State of the Global Climate 2022 Report released by the World Meteorological Organization, greenhouse gas concentrations have increased over the years, resulting in continuous heat accumulation. The past eight years were the warmest on record globally. In 2022 alone, extreme heat waves, drought, and wildfires hit different parts of the world, incurring losses of billions of US dollars. The impact of warming as a prominent feature of climate change has been unprecedented. Consequently, the climate system has become less stable, and extreme meteorological and hydrological events, such as drought and floods, occur with an increasing probability and have a lasting impact. For these reasons, monitoring and predicting droughts and floods are particularly important and have attracted widespread attention at home and abroad [4–9]. Southwest China occupies an extensive area and is susceptible to drought due to the influence of water vapor over the Bay of Bengal and the south trough. Southwest China is a high-risk region for severe drought. Great variations in the altitude, population, land use type, and soil type have given rise to the complexity and diversity of drought-inducing factors across southwest China. The causes and formative mechanisms of drought also vary dramatically in this region [10–12]. Constructing weather stations is particularly difficult in the Qinghai–Tibet Plateau and the Yunnan–Guizhou Plateau, which are high-altitude regions. The sparse distribution of weather stations in these regions exacerbates the scarcity of meteorological data. Conventional drought indices no longer apply to the entirety of southwest China and may result in large deviations when used for drought monitoring and prediction. Southwest China suffers from more severe droughts than ever in terms of global warming due to the dual impacts of natural and human activities [13–15]. Therefore, a precise and dynamic drought monitoring model for southwest China is urgently needed to reduce losses.

Drought indices offer a quantitative description of drought duration, severity, and the extent of the disaster. They constitute the foundation for modern drought monitoring and predictive modeling [16,17]. Two types of data sources are used for calculating drought indices: data at ground stations and gridded remote sensing data. Data at ground stations are obtained using calculations from real-time meteorological, agricultural, or hydrological measurements. Ground measurement data obtained using this method exhibit a higher precision but have a smaller coverage area, making them more expensive. Moreover, weather stations are sparsely distributed in some regions, and data at these stations are inadequate for describing the spatial distribution of drought. Spatial interpolation is useful for assessing drought conditions where weather stations are sparsely distributed. However, drought monitoring and prediction are usually less accurate for regions using interpolation due to the topographic complexity and uncertainty regarding the interpolation algorithm. Remote sensing data have the advantages of extensive coverage, high spatial resolution, and strong timeliness, and they can make up for the defects of weather station observation data. Along with developments in remote sensing technology, new denoising algorithms and atmospheric correction algorithms have led to many remote sensing drought indices being proposed and used for drought monitoring on a global or regional scale. Remote sensing drought indices have attracted increasing attention in recent years [18,19].

Drought adversely impacts plant growth and development, which present as varying spectral features on remote sensing images. Drought can cause changes in the physiological and biochemical parameters of plants, which further lead to spectral changes. Therefore, most remote sensing drought indices define drought by monitoring the vegetation status on the land surface. The normalized difference vegetation index (NDVI) is the most commonly

used approach. Some NDVI-related remote sensing drought indices, such as the vegetation condition index (VCI), standardized vegetation index, atmospherically resistant vegetation index, modified soil-adjusted vegetation index, and modified normalized difference vegetation index, are improvements and extensions of the NDVI and are used for drought recognition and monitoring [20–25].

In addition to optical reflection, the temperature condition index (TCI) can also be obtained using earth surface temperatures detected using the thermal channel of a satellite sensor. The TCI can be used to determine the vegetation pressure caused by dryness or excessive humidity [26]. Other remote sensing drought indices are an integration of ground reflectance and thermal properties. For example, the vegetation temperature condition index (VTCI) was developed based on the NDVI and LST products. Existing studies have shown that the VTCI performs highly accurately in identifying and classifying drought events and for near-real-time drought monitoring [27]. Sandholt et al. (2002) [28] performed empirical parameterization to establish the relationship between the NDVI and LST and proposed the temperature–vegetation dryness index (TVDI). They compared the TVDI with the distributed physical hydrological model based on the MIKE Système Hydrologique Européen code and found that the TVDI reflected water content changes on a more refined scale.

More than a hundred drought indices have been developed so far, but not a single drought index fully reflects drought features that are highly complex and occur in various forms. A single factor, such as precipitation, vegetation status, or soil moisture (SM), is generally considered and used to build conventional remote sensing drought monitoring models. These models cannot fully characterize drought. A regional drought is influenced by several factors simultaneously. Some researchers have recently integrated disaster-inducing factors with several drought indices to build a composite drought index for developing drought monitoring and prediction models [29]. Currently, three primary methodologies are employed to build a composite drought index: weight combination, multivariate joint distribution method, and ML [30]. ML stands out due to its nonlinearity, high estimation accuracy, and high generalization ability. The availability of multisource remote sensing products and data has increased due to advancements in space exploration technology and the emergence of satellite detection platforms with satellite-borne sensors. Also, the spatial and temporal resolutions have increased significantly. Remote sensing now provides information on precipitation, temperature, SM, terrestrial water storage, evaporation, snow, vegetation response, and plant functions [31,32]. Such information can be used to characterize drought from both temporal and spatial perspectives. In particular, National Aeronautics and Space Administration–Advanced Very High Resolution Radiometer and National Aeronautics and Space Administration–Moderate Resolution Imaging Spectroradiometer (NASA-MODIS) sensors provide different types of remote sensing products characterized by longer time scales and higher spatial and temporal resolutions. These remote sensing products serve as spatiotemporal Big Data for drought monitoring, often involving building a composite remote sensing index using ML methods based on data mining and a data-driven approach.

ML has developed exponentially in recent years in the field of artificial intelligence. The application scope of ML methods has expanded, and many breakthroughs have been achieved. A growing number of researchers have used a single ML algorithm or an ML approach based on ensemble learning (bagging, boosting, and stacking) to monitor and predict drought by satellite-derived drought factors in different parts of the world and on different time scales [20–22,33,34]. The ML models [21,33–40] used include RF, boosted regression tree (BRT), Cubist, support vector machine (SVM), deep forward neural network (DFNN), k-means, principal component analysis, artificial neural network (ANN), generalized additive model, deep learning (DL), classification and regression tree (CART), multivariate adaptive regression splines, flexible discriminant, convolutional neural network (CNN), XGBoost, decision tree (DT), and so forth. Park et al. (2016) [34] monitored meteorological and agricultural drought in the arid region of Arizona and New Mexico and

the humid region of North Carolina and South Carolina by incorporating 16 remote sensing drought factors from MODIS and tropical rainfall measuring mission (TRMM) satellite sensors using RF, BRT, and Cubist models. The results showed that RF produced the best performance ($R^2 = 0.93$, $RMSE = 0.3$) for standard precipitation index (SPI) prediction among the three approaches. Shen et al. (2019) [36] constructed and tested a comprehensive drought monitoring model using the DFNN based on the satellite data including MODIS and TRMM as multisource remote sensing data in the Henan Province of China. The results demonstrated that the comprehensive drought model had good applicability in monitoring meteorological and agricultural droughts. The consistency rate of drought grades between the drought indicators of the model output and the comprehensive meteorological drought index, measured at the site scale, was 85.6% and 79.8% for the training set and the test set, respectively. The CC between the model's drought index and the SPEI was between 0.772 and 0.910. Sardar et al. (2022) [38] discussed an ensemble of CNN and Barnacles Mating Optimizer (BMO) to enhance the efficiency of a CNN model for drought prediction by inputting the NDVI, soil-adjusted vegetation index, atmospherically resistant vegetation index, and enhanced vegetation index (EVI) calculated from the satellite data for the Kolar region of Karnataka. The CNN model achieved an accuracy below 90% during training and 91% during testing. However, the accuracy of the ensemble model (CNN with BMO) reached 92% during training and increased to 94% during testing. Ali et al. (2023) [41] employed XGBoost and ANN to downscale the Gravity Recovery and Climate Experiment (GRACE) satellite's terrestrial water storage anomaly (TWSA) from 1 to 0.25 for improved understanding of hydrological droughts in the Indus Basin Irrigation System. The findings revealed that the XGBoost model outperformed the ANN model with Nash–Sutcliffe Efficiency (0.99), Pearson correlation (0.99), root mean square error ($RMSE$) (5.22 mm), and mean absolute error (MAE) (2.75 mm) between the predicted and GRACE-derived TWSA. Relevant studies have shown that drought monitoring and prediction based on ML can achieve superior overall performance compared with conventional regression analysis, time-series statistical models, and physical models [42,43].

Southwest China is located in the upper reaches of the Yangtze and Pearl rivers. It serves as an important ecosafe barrier, although it is ecologically vulnerable and sensitive to climate change. Southwest China is known for its highly diversified terrain and topography because the Qinghai–Tibet Plateau (the plateau with the highest altitude and most complex terrain in the world), the Yunnan–Guizhou Plateau, the Hengduan Mountainous Region, and the Sichuan Basin are situated in this region. The climate in southwest China is influenced not only by the Qinghai–Tibet Plateau but also by the South Asian and East Asian monsoons. The formative factors of weather and climate are also highly complex in southwest China. Historically, southwest China has frequently experienced severe and high-grade droughts. Drought events have become more frequent and severe as a result of global warming [44]. Fu et al. (2022) [45] generated spatiotemporally continuous SPEI data spanning from 1901 to 2018 in southwestern China using four ML approaches (DT, RF, gradient BRT, and extra tree). The results indicated that four CART approaches could provide valid local drought information by downscaling the Estación Experimental de Aula Dei data. Mei et al. (2022) [46] proposed a novel prediction model for predicting the monthly MCI values of the representative five stations of the Yunnan Province from 1960 to 2020. This model combined the recurrent neural network (RNN) based on a gated recurrent unit and CNNs with optimization using the modified particle swarm optimization algorithm. The results showed significantly improved skills in terms of the $RMSE$ (0.301), MAE (0.237), and Nash–Sutcliffe efficiency coefficient (0.998) for predicting MCI values for the first month in the future. Currently, ML methods are less used for drought monitoring and prediction in southwest China, especially the combination of multisource remote sensing information and ML methods. The present study mainly performed the following tasks based on previous studies: (1) A MODIS LST reconstruction model was proposed based on the multisource remote sensing information and RF, and the accuracy of the model was evaluated and validated on a spatiotemporal scale with the observed data from meteorological stations in

southwest China. (2) A remote sensing drought monitoring model was constructed based on the multisource remote sensing information, and a reconstructed LST was achieved through the application of the RF model and XGBoost. The model accuracy was evaluated and validated with SPEI1 values calculated from observed data at meteorological stations in southwest China. (3) A comprehensive evaluation of the output results of the constructed remote sensing drought monitoring model was also conducted based on the monthly MCI values calculated from the observed data of meteorological stations in southwest China as well as actual disaster data for four seasons. The purpose was to improve dynamic drought monitoring and prediction in regions with complex terrain, varied topography, influential weather and climate factors, and sparse distribution of weather stations.

2. Materials and Methods

2.1. Study Areas

Southwest China encompasses five provinces and municipalities: Sichuan, Yunnan, Guizhou, Chongqing, and Tibet. This region lies between 78°42'E–110°11'E and 21°13'N–36°53'N, adjoining Bhutan, Pakistan, Nepal, India, Laos, and Burma, and covering a total area of 2.34 million km², accounting for about 24.5% of the total land area of China [47]. Widely known for its complex and diverse landforms and undulating terrain, southwest China covers the Qinghai–Tibet Plateau, Yunnan–Guizhou Plateau, Sichuan Basin, and the surrounding mountains. It belongs to the subtropical monsoon and alpine climate zones, with an annual average temperature of −2.8 to 23.9 °C and annual precipitation of 600–2300 mm. The precipitation is unevenly distributed with greater precipitation in the east than in the west. The difference can be up to five times in regions with the least and most precipitation [48,49]. The unique climate and landforms in southwest China have given rise to a large variety of vegetation. This vegetation is primarily divided into nine categories. Grassland, meadows, and alpine vegetation are mainly found in high-altitude regions; mid-altitude regions are dominated by bushes, coniferous forests, broad-leaved forests, and swaps; and low-altitude regions are occupied by grasses and cultivated vegetation [50,51].

Southwest China serves as an important ecosafe barrier, although the region itself is ecologically vulnerable and sensitive to climate change. According to historical documents and drought statistics since the founding of the People's Republic of China, droughts occur almost every year in southwest China with the occurrence cycle being 3–6 years for moderate droughts and 7–10 years for severe droughts. Extreme climate events, such as severe droughts, have become increasingly frequent due to continuous global warming in recent years in southwest China where precipitation is abundant, the climate is generally humid, and drought is not uncommon. However, drought disasters now hit southwest China more frequently and at a greater intensity than at any other time in history. Persistent droughts can exert a large impact on local social and economic development.

2.2. Data

2.2.1. Data Sources

The drought data used in the present study mainly originated from two sources: satellite remote sensing and weather station observation data. Remote sensing data included vegetation index products (MOD13A3 and MYD13A3), LST products (MOD11A2 and MYD11A2), land cover-type products (MCD12Q1), and precipitation products (TRMM3B43) from the Terra and Aqua satellites of NASA between 2010 and 2019. The temporal and spatial resolutions of vegetation index products were 1 month and 1 km × 1 km, respectively. The temporal and spatial resolutions of the LST products were 8 days and 1 km × 1 km, respectively. The temporal and spatial resolutions of land cover-type products were 1 month and 0.5 km × 0.5 km, respectively. The temporal and spatial resolutions of precipitation products were 1 h and 0.25° × 0.25° [52,53], respectively. The land use-type data originated from the Institute of Geographic Sciences and Natural Resources Research (<https://www.resdc.cn/>), accessed on 1 March 2021), Chinese

Academy of Sciences, with a spatial resolution of $1 \text{ km} \times 1 \text{ km}$. The SM products were obtained from the European Space Agency's Climate Change Initiative for Soil Moisture (<https://www.esa-soilmoisture-cci.org/>, accessed on 15 March 2021) with a temporal resolution of 1 day and a spatial resolution of $0.25^\circ \times 0.25^\circ$. The digital elevation model (DEM) data were obtained from the Geospatial Data Cloud (<https://www.gscloud.cn/search>, accessed on 3 January 2021) with a spatial resolution of $30 \text{ m} \times 30 \text{ m}$. Daily meteorological observations included temperature, precipitation, and ground temperature data recorded at a depth of 0 cm. These observations were made at 144 weather stations across southwest China spanning 1980 to 2019. The data were obtained from data.cma.cn. The geographic location of the study area and the distributions of meteorological stations and elevation are depicted in Figure 1.

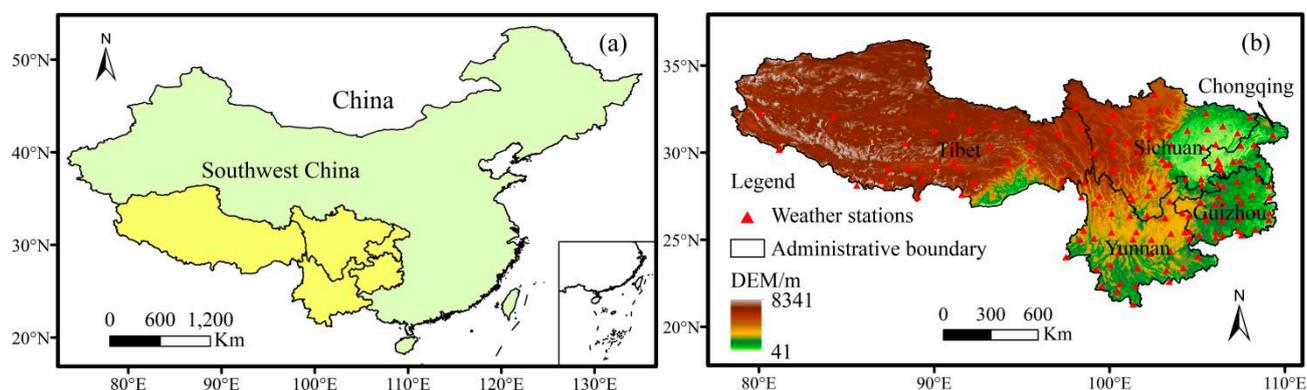


Figure 1. Geographical location (a) and spatial distribution of weather stations and DEM (b) in southwest China.

2.2.2. Data Preprocessing

The subsequent data preprocessing procedures mainly involved data format conversion, reprojection, resampling, and study area extraction because the MODIS data chosen for this study had already been subjected to atmospheric and aerosol corrections. The MODIS Reprojection Tool provided by NASA (https://lpdaac.usgs.gov/tools/modis_reprojection_tool, accessed on 20 March 2021) was first used for the projection transformation of MOD11A2, MYD11A2, MOD13A3, MYD13A3, and MCD1Q1 data. The following indices were extracted and calculated: NDVI, EVI, red band reflectance (Red_reflectance), near-infrared band reflectance (NIR_reflectance), day- and night-time surface temperature bands (LST_Day and LST_Night), and International Geosphere–Biosphere Programme (IGBP) Type 1 (according to the IGBP land cover classification system). Invalid values were removed from the images using the quality control document. A program was written to convert TRMM3B43 data and calculate monthly precipitations. The Arc Geographic Information System (ArcGIS) software was used for the projection transformation of SM, TRMM3B43, and DEM data. The slope and aspect were calculated from the DEM data. The spatial resolution of these resampling processes was set at $1 \text{ km} \times 1 \text{ km}$. As the temporal resolution of LST data was 8 days, we performed weighted pooling of all data to obtain the monthly LST, where weight was defined as the percentage of the month that each 8-day period accounted for. The temporal resolution of the SM data was set at 1 day, and the weighted pooling of the daily SM data was performed to obtain monthly data. Finally, all the remote sensing data were adjusted to match the geographical boundaries of southwest China using the ArcGIS software. Furthermore, the weather station observation data were subjected to three data preprocessing procedures for quality control: internal consistency check, climatic threshold value check, and station extreme value check. We used the means of the same meteorological element for the same day in other years for interpolation to fill some of the missing values at 144 weather stations so as to ensure the scientificity and accuracy of the meteorological observation data. We used ArcGIS and Python software

to extract pixel values of the remote sensing data for the corresponding weather stations based on the latitude and longitude of the weather stations and the time scale of data, thereby synchronizing weather station observation data and remote sensing data.

2.3. Methods

2.3.1. Calculation of Remote Sensing and Meteorological Drought Indices

Drought indices are critical for drought monitoring and prediction, serving as a prerequisite for accurate assessment. Each drought index possesses distinct features. Among remote sensing drought indices, the NDVI and EVI can be used for drought monitoring because they reflect the vegetation status. The TRMM-standardized precipitation index (TRMM-SPI) captures precipitation distribution and abnormalities in the study area. VCI is derived from the NDVI and minimizes season-related noises. The TCI is derived from LST and reflects surface sensible heat flux because stomatal closure in vegetation due to drought leads to increased sensible heat flux [54]. The NDVI contains atmospheric noises, whereas the EVI inherits the advantages of the NDVI and overcomes some defects associated with the NDVI, including sensitivity to soil background effects, oversaturation of the NDVI in areas with high-rank vegetation cover, and reliance on atmospheric correction. In addition, the EVI is more sensitive to vegetation than the NDVI and therefore increases the sensitivity of vegetation monitoring. The EVI usually performs better when applied to vegetation degradation monitoring and quantitative analysis of vegetation resources [55]. In this study, we calculated the VCI and TCI and estimated the VTCI, TVDI, and vegetation supply water index (VSWI) based on the NDVI, LST, and EVI, respectively. Both the VTCI and TVDI were calculated based on the triangular relationship between the VCI and LST. They reflected surface evapotranspiration by capturing the changes in the LST and thereby identifying drought occurrence by estimating the changes in the SM content. The VSWI was calculated from the LST and VCI and predicted drought occurrence in the case of an increase in leaf canopy temperature and a decrease in photosynthesis [54].

Among meteorological drought indices, the SPEI considers both precipitation and potential evapotranspiration based on the water balance model of the SPI. The SPEI describes the multiscale features of a drought system. It is particularly suitable for drought study in the climate change context because it is subjected to fewer restrictions in geographical and climatic conditions [16]. The MCI, developed by the National Climate Center by integrating several drought indices, has already been used in China's meteorological businesses [56]. In the present study, we used preprocessed remote sensing data and weather station observation data to calculate the VCI, TCI, $VTCI_{NDVI}$, $VTCI_{EVI}$, $TVDI_{NDVI}$, $TVDI_{EVI}$, $VSWI_{NDVI}$, $VSWI_{EVI}$, TRMM-SPI, SPEI, and MCI. The indices were used as input and output parameters for the ML model and served as a means to validate the performance of the model in assessing drought. The calculation formulas for the selected remote sensing drought indices and meteorological drought indices are presented in Table 1.

Table 1. Summary of selected drought indices.

Type	Index	Formula	Note	Reference
Drought index calculated based on remote sensing information	NDVI	$\frac{\rho_{nir}-\rho_{red}}{\rho_{nir}+\rho_{red}}$	ρ_{nir} and ρ_{red} are near-infrared and red bands, respectively.	[22]
	EVI	$G \frac{\rho_{nir}-\rho_{red}}{\rho_{nir}+C_1\rho_{red}-C_2\rho_{blue}+L}$	ρ_{blue} is the blue band; G, C ₁ , C ₂ and L are empirical coefficients.	
	TRMM-SPI	$k - \frac{c_0+c_1k+c_2k^2}{1+d_1k+d_2k^2+d_3k^3}$	SPI is calculated using the TRMM3B43 data.	[57]
	VCI	$\frac{100(NDVI-NDVI_{min})}{NDVI_{max}-NDVI_{min}}$	NDVI _{min} and NDVI _{max} are the minimum and maximum values of the NDVI, respectively; the smaller the VCI, the more likely drought will occur.	[25]
	TCI	$\frac{100(LST_{max}-LST)}{LST_{max}-LST_{min}}$	LST _{max} and LST _{min} are the maximum and minimum values of LST, respectively; the smaller the TCI, the more severe the drought may be.	
	VTCI _{NDVI}	$\frac{(a_1+b_1NDVI)-LST}{(a_1+b_1NDVI)-(a_2+b_2NDVI)}$	a ₁ , b ₁ , a ₂ , and b ₂ are the regression equation coefficients of the dry and wet edge in the relationship between the NDVI and LST, respectively. The value ranges of the four indices are all between (0,1). Among these, the smaller the value of VTCI _{NDVI} and VTCI _{EVI} , the more severe the drought may be, and the larger the value of TVDI _{NDVI} and TVDI _{EVI} , the more severe the drought may be.	[22,34,58–60]
	VTCI _{EVI}	$\frac{(a_1+b_1EVI)-LST}{(a_1+b_1EVI)-(a_2+b_2EVI)}$		
	TVDI _{NDVI}	$\frac{LST-(a_2+b_2NDVI)}{(a_1+b_1NDVI)-(a_2+b_2NDVI)}$		
	TVDI _{EVI}	$\frac{LST-(a_2+b_2EVI)}{(a_1+b_1EVI)-(a_2+b_2EVI)}$		
	VSWI _{NDVI}	$\frac{NDVI}{LST}$	The smaller the VSWI _{NDVI} and VSWI _{EVI} , the more severe the drought may be.	
VSWI _{EVI}	$\frac{EVI}{LST}$			
Drought index calculated based on meteorological station data	SPEI	$W - \frac{c_0-c_1W+c_2W^2}{1+d_1W+d_2W^2+d_3W^3}$ ($p \leq 0.5$); $-(W - \frac{c_0-c_1W+c_2W^2}{1+d_1W+d_2W^2+d_3W^3})$ ($p > 0.5$); $W = \sqrt{-2\ln(P)}$	Calculated according to the Thornthwaite method recommended by Vicente-Serrano. p is the cumulative probability, W is the cumulative probability weighted moment, and $c_0, c_1, c_2, d_1, d_2,$ and d_3 are all constant values.	[56]
	MCI	$K_a \times (a \times SPIW_{60} + b \times MI_{30} + c \times SPI_{90} + d \times SPI_{150})$	K_a is the seasonal adjustment coefficient, SPIW ₆₀ is the standardized weighted precipitation index for the past 60 days, MI ₃₀ is the relative humidity index for the past 30 days, SPI ₉₀ is the SPI for the past 90 days, SPI ₁₅₀ is the SPI for the past 150 days, and $a, b, c,$ and d are the weight coefficients of these indices.	[56,61]

The drought grades of SPEI and MCI values calculated based on the meteorological station data were consistent, as shown in Table 2. SPEI1, SPEI3, and SPEI6 represent the SPEI values of the 1-month scale, 3-month scale, and 6-month scale, respectively.

Table 2. Classification of drought grades based on SPEI and MCI values [56].

Drought Grade	Type	SPEI/MCI Value
1	No drought	$-0.5 <$
2	Mild drought	$(-1.0, -0.5]$
3	Moderate drought	$(-1.5, -1.0]$
4	Severe drought	$(-2.0, -1.5]$
5	Extreme drought	≤ -2.0

2.3.2. ML Model

(1) RF Model

The RF model was proposed by Breiman (2001) and is based on CART, which is one of the DT algorithms [62]. The rule-based ML, including decision trees and RF, has been widely used in remote sensing applications [19,34,40,63–65]. The RF method uses an ensemble approach that combines multiple decision trees to make predictions. The name “random forest” refers to data prediction accomplished using many independent DTs (a “forest”) through randomly selected training samples and variables at each node, which alleviates the well-known problems of CART such as overfitting and sensitivity to training data [66]. A randomly selected subset of training samples is used to produce a tree. The final decision from multiple trees is made by aggregating individual tree results

based on an averaging approach for regression or a majority voting for classification. The RF model calculates the increased percentage of mean square error using the out-of-bag data when a variable is permuted with a random value [62]. Based on this information, the relative importance of a variable (i.e., the contribution of the variable to predict a target variable) can be identified. Therefore, RF can decrease the variance and obtain more precise prediction results than the common tree-based algorithms.

Let D be a training dataset in an M -dimensional space X , and let Y be the class feature with the total number of c distinct classes. The construction of an RF involves a three-step process [62,66]:

Step 1: Training data sampling: Use the bagging method to generate the K subsets of training data $\{D_1, D_2, \dots, D_K\}$ by randomly sampling D with replacement;

Step 2: Feature subspace sampling and tree classifier building: For each training dataset D_i ($1 \leq i \leq K$), use a DT algorithm to grow a tree. At each node, randomly sample a subspace X_i of F features ($F \ll M$), compute all splits in subspace X_i , and select the best split as the splitting feature to generate a child node. Repeat this process until the stopping criteria are met, and a tree $H_i(D_i, X_i)$ built by training data D_i under subspace X_i is thus obtained;

Step 3: Decision aggregation: Ensemble the K trees $\{h_1(D_1, X_1), h_2(D_2, X_2), \dots, h_K(D_K, X_K)\}$ to form an RF and use the majority vote of these trees to make an ensemble classification decision.

The algorithm has two key parameters, i.e., the number of K trees to form an RF and the number of F randomly sampled features for building a DT. According to Breiman [62], parameter K can be set to 100 and parameter F can be computed by $F = \lceil \log_2 M + 1 \rceil$. For large and high-dimensional data, larger values of K and F should be used.

The LST is an important parameter characterizing land surface energy and water balance and is closely related to the SM content. It plays a crucial role in drought monitoring and prediction. In the feature space of vegetation index-LST (VI-LST), the LST value directly affects the distribution of data points on a scatterplot. This further affects the results of remote sensing inversion of drought indices, as shown in Table 1, including TCI, VTCl, TVDI, and VSWI. However, the following three problems usually exist for the MODIS LST products: (1) The MODIS sensor passes over any given region four times a day with a fixed transit time. Therefore, the sensor cannot acquire data for a given region throughout the day or at a specified time. (2) Given the large differences in climate and topography across the regions, the accuracy of MODIS LST products fluctuates significantly. (3) In the presence of cloud cover, it is the cloud top temperature rather than the LST that is actually observed [67]. We discovered during the preprocessing of MODIS LST data that the LST_Day and LST_Night data of MOD11A2 and MYD11A2 contained a large number of missing and error values, implying significant data errors.

RF in ML is an ensemble learning algorithm that uses a DT as the base learner and is considered highly accurate. An introduction of a tree ensemble endows RF with the ability to process nonlinear data and handle default values and data anomalies. Even if some of the features are already lost, RF may still ensure accuracy in monitoring and prediction. In addition, RF is relatively simple to implement and can balance the errors between datasets. Xiao et al. (2021) [68] proposed an improved LST reconstruction method for cloud-covered pixels by building a linking model for the MODIS LST with other surface variables based on an RF regression method. The validation with in situ observations revealed that the reconstructed cloud-covered LSTs performed similarly to the LSTs on clear-sky days with correlation coefficients of 0.92 and 0.89, respectively. The unbiased *RMSE* was calculated to be 2.63 K. Chen et al. (2020) [69] also proposed a novel algorithm based on RF to reconstruct MODIS LST. The experimental results indicated that the algorithm had the capacity to enhance the estimation of MODIS LST products in terms of accuracy and data availability. Sun et al. (2023) [70] used the RF model to produce a set of high-quality NDVI products to represent actual surface characteristics more accurately and naturally. Notably, the RF algorithm exhibited a *MAE* of 0.024 and a *RMSE* of 0.034, besides a R^2

value of 0.974. To effectively address the issue of significant blank areas resulting from frequent cloud cover in current thermal infrared-based LST products, Zhao et al. (2020) [71] introduced a novel method for reconstructing the cloud-covered LSTs of Terra MODIS daytime observations using the RF regression approach. They applied the RF regression approach in southwestern Europe. The reconstructed LSTs showed similar spatial patterns compared with clear-sky LSTs from temporally adjacent days, demonstrating a stable and reliable performance. Wang et al. (2020) [72] indicated that RF was superior to SVM and ANN in reconstructing and supplementing missing data.

Therefore, in the present study, we first reconstructed and supplemented the missing LST data using a built RF model and multiple impact factors related to remotely sensed LST monitoring, aiming to improve the precision of the MODIS LST products.

(2) XGBoost Model

The XGBoost model is an ensemble learning algorithm proposed by Chen and Guestrin (2016) [73]. It is an ML technique designed for regression and classification tasks. It constructs a prediction model in the form of an ensemble of weak prediction models. As an efficient and scalable variant of the gradient boosting machine, XGBoost has recently won several ML competitions based on its convenience, parallelism, and impressive predictive accuracy [74]. It is based on a gradient boosting decision tree (GBDT) and further optimizes performance by improving the model to fit the target and adding regularization terms to the objective function. The GBDT algorithm uses only the first-order derivative information for optimizing the loss function, whereas the XGBoost algorithm performs a second-order Taylor expansion of the loss function. By incorporating both the first- and second-order derivative information, XGBoost achieves a better fit to the loss function and reduces errors during the optimization process. The specific definition is provided as follows [64,75,76].

Given a training dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, $x_i \in X \subseteq \mathbb{R}^m$, $y_i \in Y \subseteq \mathbb{R}$, X is the input space, and Y is the output space. XGBoost can be expressed as an additive model as follows:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (1)$$

where \hat{y}_i represents the predicted value of the training model, f_k represents the k th sub-model; and x_i represents the i th input sample. The optimization objectives of the XGBoost algorithm include a loss function and a regularization term, and the final optimization objectives can be determined as:

$$L(t) = \sum_{i=1}^n l(y_i, \hat{y}_i(t-1) + f_t(x_i)) + H(f_t) \quad (2)$$

where $L(t)$ represents the objective function at the t th iteration; y_i represents the class-label of the original sample; $\hat{y}_i(t-1)$ represents the predicted value of the model during $t-1$ model iterations of the sample; $f_t(x_i)$ represents the predicted value of the model during the t th model iteration of the sample; and $H(f_t)$ is the regularization term of the objective function. The Taylor expansion of Equation (2) yields:

$$\tilde{L}(t) \cong \sum_{j=1}^n [\omega_j \sum_{i \in I_j} g_i + \frac{1}{2} \omega_j^2 (\sum_{i \in I_j} h_i + \lambda)] + \gamma T \quad (3)$$

where g_i represents the first-order gradient of the sample x_i ; h_i represents the second-order gradient of the sample x_i ; w_j represents the output value of the j th node; λ and γ are the coefficients of the regularization term to prevent the model from overfitting; and I_j is the subset of samples in the j th leaf node.

The training process of the XGBoost model was used to solve Equation (3) and find the best ω_j^* and the optimal solution of the corresponding objective function:

$$\omega_j^* = -\sum_{i \in I_j} g_i / (\sum_{i \in I_j} h_i + \lambda) \quad (4)$$

Equation (5) was used to measure the quality of a tree structure; the lower the value, the better the tree structure. Therefore, when the nodes in the tree were split, Equation (6) was obtained as follows:

$$\tilde{L}(t) = -\frac{1}{2} \sum_{j=1}^r ((\sum_{i \in I_j} g_i)^2 / (\sum_{i \in I_j} h_i + \lambda)) + \gamma T \quad (5)$$

$$\text{Gain} = L_{\text{left}} + L_{\text{right}} - L_{\text{father}} \quad (6)$$

If the gain value is greater than zero, the node splitting continues; otherwise, the node splitting stops.

The RF algorithm can properly handle missing and abnormal values, but it tends to neglect the correlations between the attributes, thereby affecting the regression performance. Therefore, in the present study, a remote sensing drought monitoring model was constructed based on the multisource remote sensing information (multiple remote sensing drought indices involved), and we reconstructed the LST using the RF model and XGBoost. The rationale for selecting the XGBoost model was as follows:

- (A) XGBoost belongs to the category of rule-based models. Those models are generally better suited than DL algorithms for the datasets of moderate or small size.
- (B) XGBoost models have a higher accuracy due to the introduction of second-order Taylor expansion. The base learner of XGBoost can be a DT or a linear classifier, implying higher flexibility.
- (C) XGBoost models are convenient to build in that they can attain highly optimized performance by following a standard hyperparameter search process implemented using stratified k -fold nested cross-validation (CV). Because of the regularization term, XGBoost can also be easily trained in such a way as to reduce overfitting.
- (D) XGBoost can incorporate elements of cost-sensitive learning where a cost matrix can help influence the model to produce fewer false negatives.
- (E) XGBoost supports column sampling, and it can reduce computational load and accelerate the calculation. It has been used successfully to win several ML competitions.
- (F) Previous drought studies have obtained successful results using XGBoost for predicting meteorological indicators [74,76,77].

2.3.3. CV Method

A central tenet of accuracy assessment is that the samples used for training should not also be used for evaluation. A similar concern applies to the methods for selecting the user-specified parameters required by most ML methods. The value of these parameters can affect the accuracy of the classification, and thus, the optimization of the chosen values (sometimes called tuning) is usually required [78–82]. Tuning is generally empirical, with various values for the parameters systematically evaluated, and the combination of values that generate the highest overall accuracy is assumed to be optimal [80,83]. Excluding training samples from the samples used for evaluating the candidate parameter values reduces the likelihood of overtraining and thus improves the generalization of the classifier.

CV is an approach used for exploiting training and accuracy assessment samples multiple times and thus potentially improving the reliability of the results. It can also have a better effect on small sample data. CV involves the creation of multiple partitions, potentially allowing each sample to be used multiple times for multiple purposes, with the overall aim of improving the statistical reliability of the results. Various CV methods exist, including k -fold, leave-one-out, and Monte Carlo. Classification parameter tuning via CV has been demonstrated to improve classification accuracy in remote sensing analyses [84].

The k -fold CV method involves randomly splitting the sample set into a series of equally sized folds (groups), where k indicates the number of partitions or folds the dataset is split into. For example, if a k -value of five is used, the dataset is split into five partitions. In this case, four of the partitions are used for training data, while the remaining one partition is used for validation data. The training process is repeated five times, with each iteration using a different partition as the validation set and the remaining four partitions as the training data. The average of the results is then reported [85]. Some studies showed the merits of the cross-validation methods such as k -fold CV for parameter tuning [80,82,83], and the optimal value for k should be 5 or 10 to avoid issues associated with imbalanced datasets [86]. To improve the generalization, robustness, and reliability of the constructed models, we employed a 5-fold CV method to fine-tune the models' hyperparameters in the present study.

2.3.4. Indicators of Model Accuracy Assessment

Because the evaluation indicators of regression tasks mainly focus on the differences between predicted and true values, the accuracy of the two models was evaluated by computing and comparing statistical measures based on the differences between the observed (114 weather stations) and predicted (LST and SPEI) values. These metrics included CC, RMSE, MAE, and EVS, which were calculated using Equations (7)–(10), respectively.

$$CC = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

$$EVS = 1 - \frac{Var(y_i - \hat{y}_i)}{Var(\hat{y}_i)} \quad (10)$$

where i is the data of the i th sample point, y_i refers to the true value of the sample, \hat{y}_i refers to the predicted value of the sample, and \bar{y} and $\bar{\hat{y}}$ refer to the average of the true value sample and predicted value sample, respectively. The n symbol refers to the number of samples, and Var is the sample variance. Among the five indicators, CC is used to measure the correlation between the two variables. The value of CC is between -1 and 1 ; the closer its value is to 1 or -1 , the stronger the relationship between the true value and the predicted value. The EVS is between 0 and 1 . The closer its value is to 1 , the better the model effect. The RMSE is the deviation between the predicted value and the true value. It is often used as a standard for measuring the prediction results of ML models. The value of RMSE and MAE is between 0 and ∞ ; they are two indices greater than zero, and the closer its value is to 0 , the better the model effect [65,73].

In addition, by classifying drought grades from the model-predicted and station-calculated SPEI values (Table 2), the consistency rate and omission rate were also used to evaluate the model accuracy in our study. The consistency rate is the ratio of the number of correctly classified samples to the total number of samples for a given dataset. The omission rate is the percentage of weather stations where no drought occurred based on monitoring values from the model to all of the weather stations where drought was considered to occur based on the estimated values of the drought index. The calculation formula for both cases is as follows [87,88].

$$\text{Consistency rate} = CS/O \quad (11)$$

$$\text{Omission rate} = NR/T \quad (12)$$

where CS is the number of weather stations correctly classified by drought grade (the classification of the drought index output by the model is consistent with the classification of the drought index estimated by the meteorological station according to Table 2), O is the total number of weather stations belonging to this drought grade, NR is the number of weather stations where no drought occurred based on model monitoring, but the corresponding reality is drought, and T is the total number of weather stations with actual drought conditions.

3. Results

3.1. Reconstructing the LST Using an RF Model

3.1.1. Construction of an RF Model

Given the connections between the LST and NDVI [89], we used the LST_Day and LST_Night of the LST products MOD11A2 and MYD11A2, NDVI, EVI, Red_reflectance, and NIR_reflectance of MOD13A3 and MYD13A3, and DEM data as input for the 13 feature parameters. The measured LST values at weather stations were the objects of learning. Then, an RF model was built in Python. The total dataset consisted of 1560 samples, encompassing 13 feature parameters over 120 months (2010–2019). A training set was generated by randomly selecting 70% (1092) of the samples in the dataset, whereas the remaining 30% (468) of the samples constituted the test set. The RF model with fine-tuned parameters determined by the 5-fold CV method on the training set was used to reconstruct LST values and fill in missing LST data points. The remote sensing data and ground measurements were synchronized based on time and the latitude and longitude of weather stations.

3.1.2. Evaluation and Validation of Model Accuracy

When the 5-fold CV method was used, the training set of the RF model was split into five partitions. Among these, four of the partitions were used for training data, while the remaining one partition was used for validation data. The CV score for each fold was 0.916, 0.971, 0.960, 0.967, and 0.964, respectively. The average score was 0.955, and it was close to 1. The model had good reliability and accuracy.

The RF model for reconstructing the LST was established after parameter tuning, and the optimal parameters are detailed in Table 3. We chose *RMSE*, *MAE*, *EVS*, and *CC* as the accuracy assessment indicators to verify the accuracy of the model. The accuracy of the reconstructed values on the training set and the test set is provided in Table 4. The *RMSE* values of the reconstructed values on the training and test sets were 1.172 and 2.236, respectively, and the *MAE* values were 0.847 and 1.719, respectively. Both the values were small, indicating high model accuracy. The *EVS* values were 0.901 and 0.858 on the training and test sets, respectively. Both the values approached 1, indicating an excellent model performance. In addition, the value of *CC* was more than 0.9, and a significant correlation was observed. As discussed earlier, the LST values reconstructed by the RF model differed little from the measured ones at the weather stations. The LST values in the two groups showed a strong correlation. Thus, the reconstruction of the LST values using the RF algorithm did improve the accuracy of MODIS LST products.

Table 3. Optimal parameters of LST reconstruction using the RF model.

Parameter	Meaning	Optimal Parameter
max_features	Maximum number of features used by a single decision tree	auto
max_depth	Maximum depth of the tree	15
min_samples_split	Minimum number of samples required to split a node	2
min_samples_leaf	Minimum number of samples contained in each leaf node	1
n_estimators	Number of decision trees to build	537
bootstrap	With or without put-back sampling	True
criterion	Evaluation criteria for segmentation quality	mse

Table 4. Accuracy evaluation indicators of RF model reconstruction.

Accuracy Assessment Indicator	Training Set	Testing Set
RMSE	1.172	2.236
MAE	0.847	1.719
EVS	0.901	0.858
CC	0.944 **	0.908 **

** indicates that the result is significant at the 0.01 level.

Figure 2 shows the multi-year average spatial distributions of remote sensing data, weather station observation data, and RF-reconstructed values of LST in southwest China from 2010 to 2019. The spatial distribution of RF-reconstructed LST values agreed well with that of remote sensing data and weather station observation data. In addition, the reconstructed LST values were closer to the weather station observation data. The reason for larger differences relative to weather station observation data might be attributed to the scarcity of weather stations, especially in the Qinghai–Tibet Plateau located in western Tibet. This resulted in larger errors following spatial interpolation. Figure 3 shows the monthly variations in remote sensing data, weather station observation data, and RF-reconstructed values of the LST. When averaging data for the entire southwest China region, larger differences were observed between remote sensing data and weather station observation data due to the accuracy errors in remote sensing. On the contrary, the RF-reconstructed values were closer to weather station observation data, which indicated the applicability of the RF-based reconstruction method. A combined analysis of Figures 2 and 3 shows that the RF algorithm could reconstruct LST values over a broad range and with high resolution.

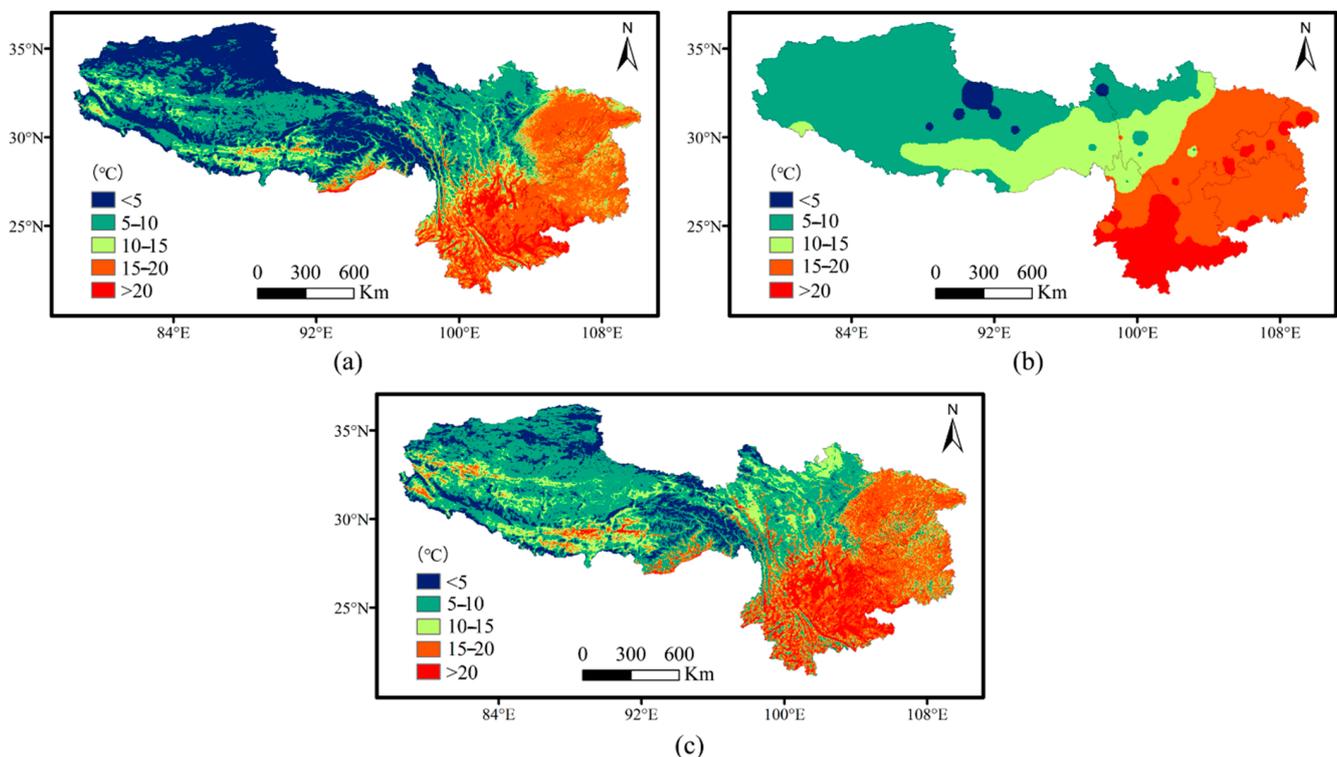


Figure 2. Multi-year average spatial distributions in the LST on remote sensing monitoring (a), meteorological station measurement (b), and RF reconstruction (c) in southwest China from 2010 to 2019.

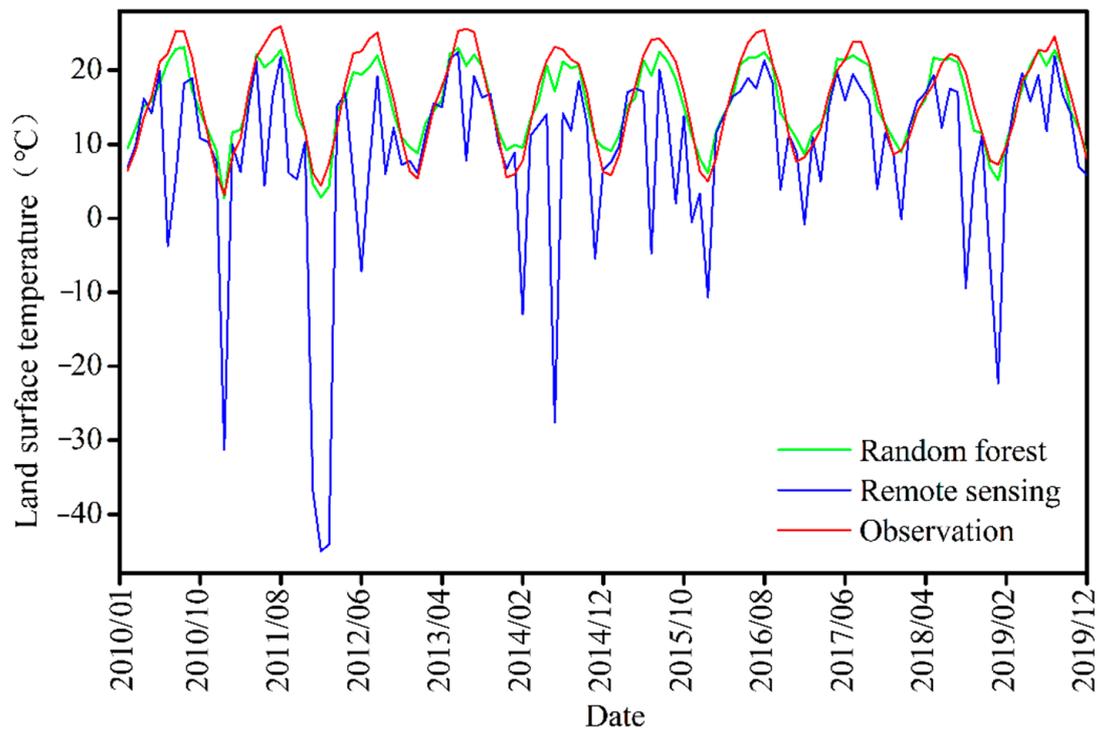


Figure 3. Monthly change in the LST on remote sensing monitoring, meteorological station measurement, and RF reconstruction in southwest China from 2010 to 2019.

3.2. Remote Sensing-Based Drought Monitoring Using XGBoost

3.2.1. Selection of Input and Output Parameters

The occurrence of drought involves a complex mechanism between various disaster-inducing factors. Factors such as vegetation cover type, DEM, and SM content exert varying impacts on drought occurrence in the study area, and reduced SM content is one of the direct causes of drought. Remote sensing drought indices also reflect drought occurrence and development by capturing the vegetation status and LST. In the present study, we fully considered various drought-inducing factors during the remote sensing monitoring of drought in southwest China, including vegetation, LST (constructed by RF model), precipitation, vegetation cover type, SM content, and DEM elevation. All these factors were considered as input parameters of the XGBoost model. In our preliminary assessment of the suitability of SPEI and MCI for drought monitoring in southwest China, we observed that the SPEI outperformed the MCI, particularly in Tibet, where weather stations were sparsely distributed [90,91]. In addition, the SPEI allowed for effective monitoring of the major drought-prone regions in each season and therefore had higher overall applicability in southwest China. Given this advantage, the SPEI was considered as the object of learning in the XGBoost model, and the predicted values were the outputs. Moreover, some input parameters might respond more slowly compared with the SPEI over the same period, resulting in synchronization issues. Furthermore, SPEI values over different time scales represent different types of drought. Therefore, we performed a correlation analysis between all input parameters and SPEI values across different time scales (1, 3, and 6 months). The analysis results are presented in Table 5. The drought severity, as indicated by the input parameters, closely matched the conditions represented by the SPEI value on a 1-month time scale. Therefore, SPEI1 was considered as the expected output parameter when constructing the drought-monitoring model.

Table 5. Correlation coefficients between input parameters and SPEI values at different time scales.

Drought Index	SM	TRMM-SPI	VCI	TCI	VTCI _{NDVI}	VTCI _{EVI}	TVDI _{NDVI}	TVDI _{EVI}	VSWI _{NDVI}	VSWI _{EVI}
SPEI1	0.488 **	0.835 **	0.101	0.105	0.044	0.066 *	−0.097	−0.071	0.187	0.096 *
SPEI3	0.499 **	0.568 **	0.096	0.090	0.021	0.004	−0.083	−0.064	0.105	0.078
SPEI6	0.471 **	0.414 **	0.110	0.067	0.080	0.062	−0.057	−0.034	0.128	0.088

Note: ** and * indicate that the results are significant at the 0.01 and 0.05 levels, respectively.

3.2.2. Building a Remote Sensing-Based Drought Monitoring Model Using XGBoost

In this study, the input parameters of the XGBoost model included all of the remote sensing drought indices in southwest China from 2010 to 2019 (Tables 1 and 5) along with land cover-type products MCD12Q1 and DEM. SPEI1 estimated from weather station observation data was the expected output parameter. The total data sample (1440) consisted of the 120 months (2010–2019) of 12 feature parameter data. The training set was generated by randomly sampling 70% (1008) of the samples, while the remaining 30% (432) of the sample constituted the test set. An XGBoost-based remote sensing drought monitoring model was built in Python. The land cover-type data were combined and classified into six main types of land use: arable land, forest land, grassland, water bodies, urban construction land, and unused land. The optimal parameters of the XGBoost model were determined by fine-tuning using a 5-fold CV method within the training set (Table 6). The CV score for each fold was 0.850, 0.927, 0.962, 0.965, and 0.949, respectively, and the average score was 0.931.

Table 6. Optimal parameters of the XGBoost regression monitoring model.

Parameter	Meaning	Optimal Parameter
n_estimators	Number of submodels	203
max_depth	Maximum depth of the tree	2
learning_rate	Learning rate of the resulting model at each iteration	0.06
min_child_weight	Minimum number of samples contained in each leaf node	2
Subsample	Proportion of random sampling	0.8
Gamma	Controls whether to post-prune	0
colsample_bytree	Controls the proportion of each random sampling column	1
colsample_bylevel	Proportion of column sampling for each node splitting in each tree	0.5
reg_alpha	Weight of the L1 regularization term	0.01
eval_metric	Measures the validation data	mse

3.2.3. Model Accuracy Evaluation

Similarly, the *RMSE*, *MAE*, *EVS*, and *CC* were chosen as the accuracy assessment indicators of the model. The results are depicted in Table 7. The *RMSE* of the monitoring values on the training and test sets was 0.135 and 0.435, respectively, and the *MAE* was 0.095 and 0.328, respectively. Both values were small, indicating a high level of accuracy for the model. The *EVS* values of the monitoring values on the training and test sets were 0.976 and 0.782, respectively. Although the *EVS* was lower on the test set, it still was more than 0.75. In addition, the *CC* of the monitoring values relative to SPEI1 was 0.982 on the training set and 0.868 on the test set, indicating significant correlation and high model monitoring performance.

Table 7. Accuracy evaluation indicators of the XGBoost monitoring model.

Accuracy Assessment Indicator	Training Set	Testing Set
RMSE	0.135	0.435
MAE	0.095	0.328
EVS	0.976	0.782
CC	0.982 **	0.868 **

** indicates that the result is significant at the 0.01 level.

According to the drought grade classification based on the SPEI (Table 2), we counted the number of stations and the consistency rate of SPEI1 values between those calculated based on data from 144 weather stations and those predicted based on the output of the monitoring model for different drought grades spanning from 2010 to 2019. The results are summarized in Table 8. The overall consistency rate of the drought grades between the two sets of data was more than 85%. The consistency rate was higher for no drought, which was 96.87%. The consistency rate was more than 70% for mild and moderate drought, and it was 62.79% for severe drought. It was the lowest for extreme drought, which was only 26.84%. The aforementioned results might be explained by the small number of extreme drought samples during the 10-year study period. Overall, the monitoring of SPEI1 values from the model agreed well with the SPEI1 values estimated by weather stations.

Table 8. Number of weather stations under each drought grade and their consistency rates from 2010 to 2019.

Drought Grade	Number of Weather Stations with Consistent Drought Grade/Total Number of Weather Stations under Each Drought Grade	Consistency Rate (%)
Extreme drought	91/339	26.84
Severe drought	712/1134	62.79
Moderate drought	1512/2077	72.80
Mild drought	1861/2624	70.92
No drought	10,624/11,106	96.87
Total	14,852/17,280	85.65

The MCI value considers the impact of precipitation and evapotranspiration over different time intervals of 30, 60, 90, and 150 days. The MCI considers more influencing factors compared with the SPEI. We also calculated MCI values at 144 weather stations in southwest China from 2010 to 2019 to further validate the accuracy of the monitoring values. The consistency rate of monthly drought grades was determined using model monitoring output and MCI values estimated by 144 weather stations. The omission rate was calculated as well (Figure 4). The overall consistency rate between drought grades identified using model monitoring and MCI values was 67.88%, indicating satisfactory consistency. The overall omission rate was 18.89%. This result demonstrated the model's ability in drought monitoring. The consistency rates consistently exceeded 58% across all months, with the highest rate observed in September at 75.07%. The consistency rate was the lowest in February at 58.26%. However, the omission rate in February was also lower at 15.58%. The consistency rate was the lowest in winter and higher in the other three seasons, which might be due to the selection of K_a values during MCI calculation. The K_a value is usually determined based on the sensitivity of main crops to SM in different growth and development stages in different regions and seasons. While its typical range is between 0.4 and 1.2, for areas without crops or vegetation growth and perennial dry areas, the K_a value is considered as 0 [56]. Hence, the applicability of the MCI in southwest China varied with the variation in weight coefficients and K_a values involved in MCI calculation. This further affected the consistency rate between the monitoring and MCI values [92].

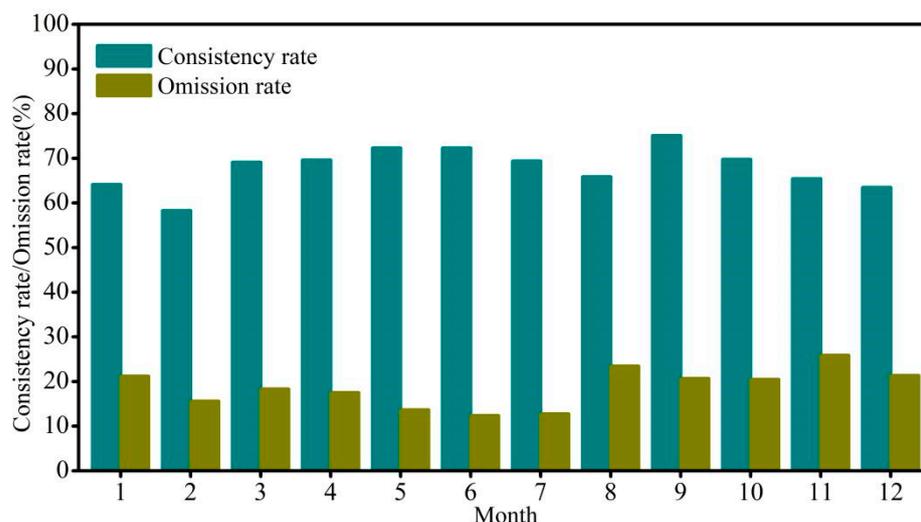


Figure 4. Consistency and omission rates of drought each month between model monitoring and MCI values estimated by 144 weather stations.

Additionally, based on historical drought records, severe droughts occurred in various seasons in southwest China during 2010–2019: during the spring and summer of 2010, the autumn of 2011, and the winter of 2012. A more detailed overview about the main droughts is presented in Table 9 [93–95]. We used the drought monitoring model in four seasons, such as during March and June of 2010, September of 2011, and February of 2012, to conduct a comprehensive validation of the monitoring performance of the model. The spatial distributions of drought grades identified using monitoring values from the model and MCI and SPEI values estimated at 144 weather stations are illustrated in Figure 5.

Table 9. Historical actual drought conditions for the four seasons selected from 2010 to 2012.

Season \ Region	Sichuan	Chongqing	Yunnan	Guizhou	Tibet
March 2010	Mild-to-severe drought in the southern region during the first 10-day period, with the drought relieved during the third 10-day period	No apparent drought	Mild-to-severe drought in the southern region and extreme drought locally in the northern region during the first 10-day period with the drought relieved during the third 10-day period	Mild-to-severe drought in the southern region and extreme drought in the southwestern region during the first 10-day period with the drought relieved during the third 10-day period	Mild-to-severe drought in the central region with an extreme drought locally during the first 10-day period with the drought continuing into the third 10-day period
June 2010	No apparent drought	No apparent drought	Moderate drought in the central region during the second 10-day period and a mild drought in the central and northern regions during the third 10-day period	No apparent drought	Mild-to-extreme drought in the central region, with an extreme drought mainly occurring near Nyima County of Nagqu City
September 2011	Drought of moderate severity and above in the southeastern region, with a severe drought locally	Drought of moderate severity and above in the southwestern region, with a severe drought locally	Drought of moderate severity and above in the northeastern region with a severe drought locally	Drought of moderate severity and above in most regions with a severe drought in northwestern and eastern regions	Moderate-to-severe drought in central and eastern regions
February 2012	Mild drought in the southwestern region during the first 10-day period with a moderate drought locally, and a severe drought in the central and western and southern regions during the third 10-day period	Moderate-to-severe drought in the central and northern regions	Mild drought in the western region during the first 10-day period, and a moderate-to-severe drought in most parts during the second 10-day period, with an extreme drought locally in the western region	No apparent drought	Mild-to-moderate drought in central and southern regions

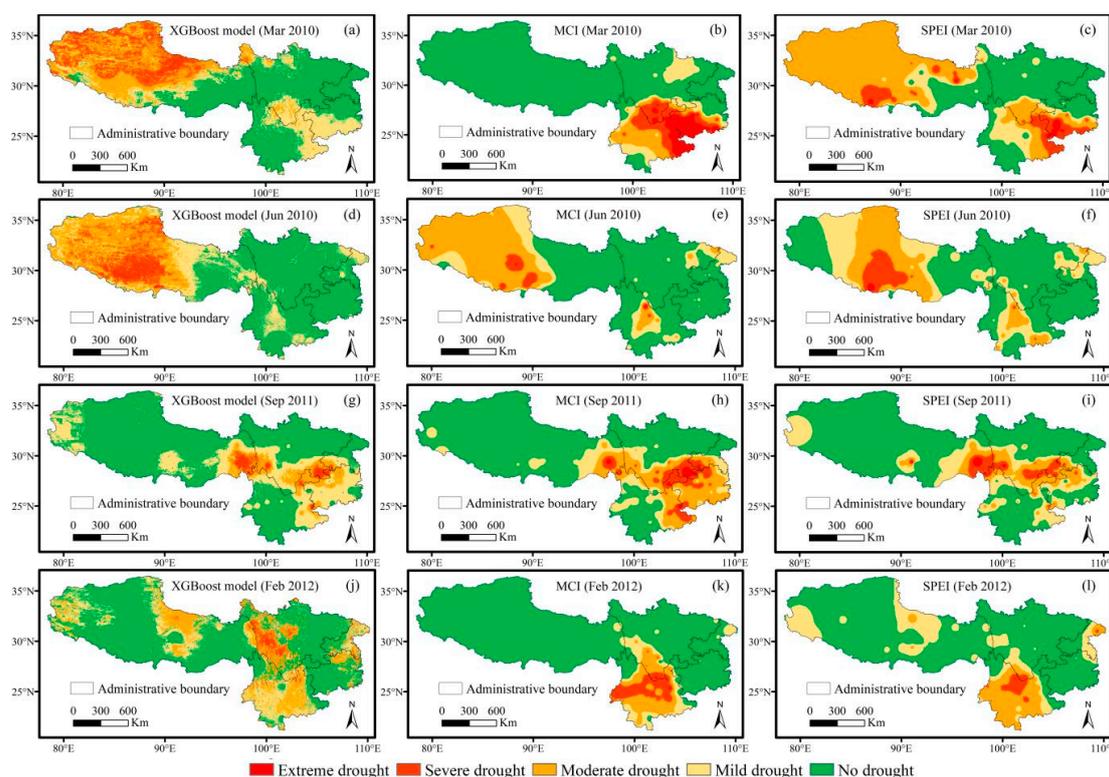


Figure 5. Spatial distribution of drought grades as determined by the XGBoost model monitoring, MCI, and SPEI in southwestern China. (a) XGBoost model (March 2010), (b) MCI (March 2010), (c) SPEI (March 2010), (d) XGBoost model (June 2010), (e) MCI (June 2010), (f) SPEI (June 2010), (g) XGBoost model (September 2011), (h) MCI (September 2011), (i) SPEI (September 2011), (j) XGBoost model (February 2012), (k) MCI (February 2012), (l) SPEI (February 2012).

A combined analysis of the findings listed in Table 9 and depicted in Figure 5 revealed that the model monitored a moderate-to-severe drought in central and western Tibet and a mild-to-moderate drought in southern Sichuan, northern and northeastern Yunnan, and southern Guizhou in March 2010. These monitoring results agreed with the actual drought conditions. A comparison of the results of the spatial interpolation of the MCI and SPEI showed that the drought severity was overestimated, which was divergent from the actual drought situation in southeastern China during the last 10 days of March. In addition, the MCI failed to ensure the effective monitoring of drought in Tibet. The model monitored drought occurrence in Yunnan and Tibet in June 2010 with a mild drought in northern–central Yunnan and a mild-to-extreme drought in Tibet. The regions affected by an extreme drought were mainly located in Nagqu and Shigatse. The model monitoring results agreed well with the actual drought situation and corresponded to the spatial interpolation of MCI and SPEI values. All three methods achieved better monitoring performance for drought in the summer in southwest China. Nevertheless, the drought grades monitored using the model were in line with the actual drought situation during the last 10 days of June. In September 2011, the model monitored a moderate-to-severe drought in southern Sichuan, northeastern Yunnan, most parts of Guizhou, southern Chongqing, and eastern Tibet. Severe droughts mainly occurred in Sichuan and Tibet as well as the junction between Yunnan, Guizhou, Sichuan, and Chongqing. A mild drought was monitored in some parts of central and western Tibet. These findings agreed well with the actual drought situations. However, the MCI and SPEI values indicated the occurrence of extreme drought, which was an overestimation compared with the actual drought situation. In addition, the SPEI exhibited a weaker monitoring performance for drought in Guizhou. Despite this disagreement, the three methods predicted similar spatial distributions of drought, and all exhibited satisfactory monitoring performance in terms of locating drought-stricken regions.

In February 2012, the model monitored a mild-to-moderate drought in central and western Tibet and a moderate-to-severe drought in central and southwestern Sichuan, most parts of Yunnan, and central–northern Chongqing. Basically, no drought was observed in most parts of Guizhou. The areas affected by a severe drought were mainly located in western and northeastern Yunnan and central–western and southern Sichuan. However, the model underestimated the severity of the drought in Yunnan compared with the actual drought situation and failed to monitor an extreme drought in western Yunnan. The MCI failed to monitor the winter drought in Tibet due to the improper value of $K_q = 0$. Both the MCI and SPEI failed to monitor the severe drought in Sichuan. Errors were found in the monitored spatial distribution of drought-affected regions. Hence, the comparison of the monitoring results using XGBoost, MCI, and SPEI with the actual drought situation suggested that XGBoost monitored similar drought grades and spatial distribution of drought in different seasons. Compared with the spatial interpolation of the MCI and SPEI at the weather stations, XGBoost achieved more refined regional monitoring. XGBoost could also monitor drought in the Qinghai–Tibet Plateau where the weather stations were sparsely distributed. However, XGBoost was less effective in monitoring an extreme drought due to the scarcity of learning samples.

4. Discussion

This study showed that the *RMSE* and *MAE* values were small for the reconstructed LSTs using the RF model based on the multisource remote sensing data from 2010 to 2019 on the training and test sets. The *EVS* was 0.901 and 0.858, respectively, approaching 1. The *CC* value was consistently more than 0.9, indicating a significant correlation. This implied that the RF model could dramatically improve the accuracy and integrity of remotely sensed LSTs. The reconstructed LSTs exhibited a spatial distribution that closely resembled the monitored values obtained by remote sensing inversion. The reconstructed LSTs demonstrated a closer alignment with weather station observation data. In addition, the RF model achieved a high performance for the Tibet region where weather stations were sparsely distributed, thereby improving the calculation accuracy for remote sensing drought indices related to LST. The aforementioned findings agreed well with those of Cheng et al. (2020) [69] and Cheng (2020) [96].

The *RMSE* and *MAE* values were small for SPEI1 values obtained using the XGBoost model based on the multisource remote sensing data and those calculated from weather station observation data on both the training and test sets, while the *EVC* and *CC* values were high. For the drought grade, the consistency rate between the two sets of values was 85%. It was 96.8% for no drought and more than 70% for mild and moderate drought. The SPEI1 values from the XGBoost model were consistent with those from the MCI values calculated from weather station observation data. The consistency rate was the highest in September and the lowest in February. Moreover, the aforementioned values well reflected the historical spatial distributions of drought and drought grades in the four seasons, such as spring of 2010 (March), summer of 2010 (June), autumn of 2011 (September), and winter of 2012 (February). Compared with the SPEI and MCI values calculated using the weather station observation data, the drought spatial monitoring using the XGBoost model was more accurate and covered more extensive areas. The drought grades monitored using the XGBoost model were closer to the actual drought situation. More importantly, the XGBoost model effectively monitored the drought situation in the Qinghai–Tibet Plateau and other parts of Tibet, where the weather stations were sparsely distributed. The monitoring results had a higher spatial resolution than the interpolated values of the MCI and SPEI. However, the consistency rate between the monitoring of a severe drought based on the SPEI1 values from the XGBoost model and those calculated from weather station observation data was 62.79%. The consistency rate between the monitoring of extreme drought was even lower at 26.84%. This result might be explained by the scarcity of severe and extreme drought samples selected from the 10-year period. Meanwhile, the overall consistency rate between the drought grades monitored using the XGBoost model and those based

on the MCI values was relatively low at 67.88%. Shen et al. (2017) [87] reported that the consistency rate between the drought grades constructed using the RF model based on remote sensing data and the comprehensive meteorological drought index for model performance validation was 74.9% for Henan. According to Jia and Zhang (2018) [97], severe and extreme drought events in southwest China in the summer and autumn of 2011 were not effectively monitored. This was because of the following reasons: (1) The results varied if the ML model, input variables, and region of interest selected varied. (2) SPEI values were the outputs from the XGBoost model, and their calculation principle diverged from that for MCI values. (3) The K_a values in MCI calculation might result in less drought monitoring for some months in Tibet with $K_a = 0$ in January to March and November to December [56]. This explained the low consistency rate between drought grades from drought monitoring and those based on MCI values in Tibet. In addition, drought severity seemed to be overestimated based on the MCI values [98]. (4) The Thornthwaite method was used to estimate potential evapotranspiration when SPEI values were determined using the XGBoost model. As the Thornthwaite method only considered temperature and precipitation factors but not wind speed, water vapor pressure, soil heat flux, and surface net radiation, the SPEI values calculated using the Thornthwaite method might differ from the real situation. Moreover, model learning might amplify such errors, resulting in significant deviations [22,56]. (5) The downscaling method for SM and TRMM data and the precision of MODIS data might also cause deviations and uncertainty, further affecting the monitoring performance of the model.

Additionally, the RF-reconstructed LST model and the XGBoost remote sensing monitoring model both used the measured LST of 144 ground meteorological stations and the SPEI1 value calculated from the measured data of the meteorological stations as the expected output parameters. When evaluating and validating the model accuracy, they were also compared and calculated with the measured values at ground meteorological stations. The distribution of meteorological stations was uneven in southwest China, with a significant concentration in the east and less concentration in the west (Fig. 1), especially in western Tibet, where meteorological stations were sparsely distributed and had high altitude. Although DEM was considered as an input parameter to the model, the lack of learning samples and measured assessment data could lead to large uncertainties in the model results, especially in western southwest China. Zhao et al. (2020) [71], in evaluating the reconstructed LST using the RF regression approach model performance, not only used LST data derived based on in situ air temperature measurements but also referred to the Global Land Data Assimilation System (GLDAS) Noah 0.25° 3 h LST data for comprehensive analysis. To enhance the model's reliability and reduce uncertainty, the high spatial and temporal resolution reanalysis data and other remote sensing data that have undergone rigorous accuracy evaluation can also be used for comprehensive analyses in the future.

5. Summary and Conclusions

Drought results from various interacting factors. The following factors make precise and accurate drought monitoring and prediction difficult: randomness, nonlinearity, and nonstationarity of drought-influencing variables; complex physical, chemical, and biological processes within the drought system; and variations in complex influencing factors across the regions, including soil texture, terrain, land management, human activities, and climatic conditions. Particularly, drought monitoring and prediction might be highly challenging in regions with complex terrain, topography, and formative factors of weather and climate and sparsely distributed weather stations. With continuous progress in remote sensing technology, multisource remote sensing data from various satellite sensors can dramatically enrich the sources of drought-related information. Multisource remote sensing data not only improve the spatial and temporal resolutions of the drought monitoring and prediction model but also provide multiple sources of long time series for real-time and dynamic drought monitoring and prediction. In the context of rapidly developing Big Data technology and artificial intelligence, multisource remote sensing data and data-driven ML

methods can be used to mine drought features from different perspectives. This method can improve the generalization ability and accuracy of the drought monitoring and prediction models. This study focused on southwest China, where droughts occur frequently and with high intensity. We calculated several drought indices based on multisource satellite remote sensing data and weather station observation data. Multiple data sources were combined for drought monitoring and prediction using two different ML methods in southwest China. The models' performance was assessed and validated using a 5-fold CV technique to assess their accuracy and reliability. Furthermore, we compared the model-generated results for monitoring spatial and temporal drought patterns, drought grades, and the extent of drought impact with historical drought records. The comparisons were made against the SPEI and MCI values estimated using data from weather station observations.

The results of this study showed that the *RMSE* of the LST values reconstructed by the RF model on the training and test sets was 1.172 and 2.236, the *MAE* was 0.847 and 1.719, and the *EVS* was 0.901 and 0.858, respectively. Furthermore, the *CC* values were all more than 0.9. The *RMSE* values of the monitoring values by the XGBoost model on the training and test sets were 0.135 and 0.435, the *MAE* values were 0.095 and 0.328, the *EVS* values were 0.976 and 0.782, and the *CC* value was 0.982 and 0.868, respectively. The consistency rate between drought grades identified using monitoring from the XGBoost model and SPEI1 values estimated at 144 weather stations was more than 85%. The consistency rate was higher for no drought, which was 96.87%; in addition, it was more than 70% for mild and moderate drought, and it was 62.79% for severe drought. The drought grades determined through model monitoring and MCI values had an overall consistency rate of 67.88% with an overall omission rate of 18.89%. The consistency rates were more than 58% in any month with the highest occurring in September, which was 75.07%. The consistency rate was the lowest in February, which was 58.26%. However, the omission rate in February was also lower, which was 15.58%. The RF-based LST reconstructive model and XGBoost model achieved high comprehensive performance, accuracy, and applicability.

Based on the result analysis and discussion, we can further increase the number of samples and input variables for ML. Potential evapotranspiration can be estimated using the Food and Agriculture Organization of the United Nations Penman Monteith method so as to calculate the SPEI. The K_n , the seasonal adjustment factor, and the weight factor are adjusted for different provinces and municipalities during MCI estimation, especially in Tibet. The spatial resolution of different sources of remote sensing data can be improved by an appropriate statistical and dynamic downscaling method, such as that based on geographically weighted regression, to reduce uncertainty. A number of different ML models were selected for comparative analysis. In addition, some swarm intelligence methods for parameter optimization can be combined with ML models representing various spatial and temporal features to build a hybrid model. The aforementioned methods are expected to raise the level of drought monitoring and prediction in southwest China and achieve higher accuracy, robustness, and generalization ability for the constructed model.

Author Contributions: X.L.: Conceptualization, data curation, visualization, writing—original draft, funding acquisition, supervision, writing—review; H.J.: data curation, investigation, software, code, writing—original draft; L.W.: conceptualization, supervision, writing—review. All authors have read and agreed to this version of the manuscript.

Funding: This study was jointly supported by the Key Research and Development (R&D) Project of the Department of Science and Technology of Yunnan Province (202203AC100005 and 202203AC100006) and the Drought Meteorological Science Research Fund of the China Meteorological Administration (IAM202201).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

- Lemos, M.C.; Eakin, H.; Dilling, L.; Worl, J. Social Sciences, Weather, and Climate Change. *Meteorol. Monogr.* **2019**, *59*, 26.1–2625. [[CrossRef](#)]
- Vicente-Serrano, S.M.; Quiring, S.M.; Pena-Gallardo, M.; Yuan, S.S.; Dominguez-Castro, F. A Review of Environmental Droughts: Increased Risk under Global Warming? *Earth-Sci. Rev.* **2020**, *201*, 102953. [[CrossRef](#)]
- Masson-Delmotte, V.; Zhai, P.M.; Pirani, A.; Connors, S.L.; Péan, C.; Berger, S.; Huang, M.T.; Yelekçi, O.; Yu, R.; Zhou, B.Q. Climate Change 2021: The Physical Science Basis. In *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; IPCC: Geneva, Switzerland, 2021; Volume 2.
- Dai, A.G. Drought under Global Warming: A Review. *Wiley Interdiscip. Rev. Clim. Chang.* **2011**, *2*, 45–65. [[CrossRef](#)]
- Yuan, X.C.; Tang, B.J.; Wei, Y.M.; Liang, X.J.; Yu, H.; Jin, J.L. China's Regional Drought Risk under Climate Change: A Two-stage Process Assessment Approach. *Nat. Hazards* **2015**, *76*, 667–684. [[CrossRef](#)]
- Jin, J.L.; Song, Z.Z.; Cui, Y.; Zhou, Y.L.; Jiang, S.M.; He, J. Research Progress on the Key Technologies of Drought Risk Assessment and Control. *Shuili Xuebao* **2016**, *47*, 398–412.
- Huang, J.P.; Chen, W.; Wen, Z.P.; Zhang, G.J.; Li, Z.X.; Zuo, Z.Y.; Zhao, Q.Y. Review of Chinese Atmospheric Science Research over the Past 70 Years: Climate and climate change. *Sci. China Earth Sci.* **2019**, *49*, 1514–1550. [[CrossRef](#)]
- Danandeh, M.A.; Rikhtehgar, G.A.; Yaseen, Z.M.; Sorman, A.U.; Abualigah, L. A Novel Intelligent Deep Learning Predictive Model for Meteorological Drought Forecasting. *J. Ambient Intell. Humaniz. Comput.* **2022**, *14*, 10441–10455. [[CrossRef](#)]
- Rahimi, B.S. Monitoring of Hydrological Drought in Khazar Basin. *Watershed Eng. Manag.* **2023**. [[CrossRef](#)]
- Zhang, Q.; Yao, Y.B.; Li, Y.H.; Huang, J.P.; Ma, Z.G.; Wang, Z.L.; Wang, S.P.; Wang, Y.; Zhang, Y. Progress and Prospect on the Study of Causes and Variation Regularity of Droughts in China. *Acta Meteorol. Sin.* **2020**, *78*, 500–521. [[CrossRef](#)]
- Wang, Y.S.; Xiao, T.G.; Dong, X.F. Characteristics of Long-Cycle Abrupt Drought-Flood Alternations in Southwest China and Atmospheric Circulation in Summer from 1961, to 2019. *Plateau Meteorol.* **2021**, *40*, 760–772.
- Huan, D.B.; Fan, K.; Xu, Z.Q. Strengthened Relationship between Summer Barents Sea Ice and Autumn Southwest China Drought after the Mid-and Late-1990s. *Trans. Atmos. Sci.* **2022**, *45*, 167–178.
- Yao, Y.B.; Zhang, Q.; Wang, J.S.; Shang, J.L.; Wang, Y.; Shi, J.; Han, L.Y. The Response of Drought to Climate Warming in Southwest in China. *Ecol. Environ. Sci.* **2014**, *23*, 1409–1417.
- Yao, Y.B.; Zhang, Q.; Wang, J.S.; Shang, J.L.; Wang, Y.; Shi, J.; Han, L.Y. Temporal-spatial Abnormity of Drought for Climate Warming in Southwest China. *Resour. Sci.* **2015**, *37*, 1774–1784.
- Sun, Z.X.; Zhang, Q.; Sun, R.; Deng, B. Characteristics of the Extreme High Temperature and Drought and Their Main Impacts in Southwestern China of 2022. *J. Arid Meteorol.* **2022**, *40*, 764–770.
- Vicente-Serrano, S.M.; Beguería, S.; López-Moreno, J.I. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *J. Clim.* **2010**, *23*, 1696–1718. [[CrossRef](#)]
- Mi, Q.C. Construction and Prediction of the Ensemble Drought Index. Master's Thesis, Shenyang Agricultural University, Shenyang, China, 2022.
- Guo, N.; Wang, X.P.; Wang, L.; Wang, L.J.; Hu, D.; Sha, S. Review of Drought Monitoring based on Remote Sensing Technology. *Adv. Meteorol. Sci. Technol.* **2020**, *10*, 10–20.
- Han, H.Z.; Bai, J.J.; Yan, J.W.; Yang, H.Y.; Ma, G. A Combined Drought Monitoring Index based on Multi-sensor Remote Sensing Data and Machine Learning. *Geocarto Int.* **2021**, *36*, 1161–1177. [[CrossRef](#)]
- West, H.; Quinn, N.; Horswell, M. Remote Sensing for Drought Monitoring & Impact Assessment: Progress, Past Challenges and Future Opportunities. *Remote Sens. Environ.* **2019**, *232*, 111291.
- Jiao, W.; Wang, L.; McCabe, M.F. Multi-sensor Remote Sensing for Drought Characterization: Current Status, Opportunities and a Roadmap for the Future. *Remote Sens. Environ.* **2021**, *256*, 112313. [[CrossRef](#)]
- Qin, Q.M.; Wu, Z.H.; Zhang, T.Y.; Sagan, V.; Zhang, Z.X.; Zhang, Y.; Zhang, C.Y.; Ren, H.Z.; Sun, Y.H.; Xu, W.; et al. Optical and Thermal Remote Sensing for Monitoring Agricultural Drought. *Remote Sens.* **2021**, *13*, 5092. [[CrossRef](#)]
- Son, B.; Im, J.; Park, S.; Lee, J. Satellite-based Drought Forecasting: Research Trends, Challenges, and Future Directions. *Korean J. Remote Sens.* **2021**, *37*, 815–831.
- Li, Z. Drought Characteristics and Prediction Models in Northeast China. Ph.D. Thesis, Shenyang Agricultural University, Shenyang, China, 2021.
- Mullapudi, A.; Vibhute, A.D.; Mali, S.; Patil, C.H. A Review of Agricultural Drought Assessment with Remote Sensing Data: Methods, Issues, Challenges and Opportunities. *Appl. Geomat.* **2023**, *15*, 1–13. [[CrossRef](#)]
- Kogan, F.N. Application of Vegetation Index and Brightness Temperature for Drought Detection. *Adv. Space Res.* **1995**, *15*, 91–100. [[CrossRef](#)]
- Wan, Z.; Zhang, Y.; Zhang, Q.; Li, Z.L. Quality Assessment and Validation of the MODIS Global Land Surface Temperature. *Int. J. Remote Sens.* **2004**, *25*, 261–274. [[CrossRef](#)]
- Sandholt, I.; Rasmussen, K.; Andersen, J. A Simple Interpretation of the Surface Temperature/Vegetation Index Space for Assessment of Surface Moisture Status. *Remote Sens. Environ.* **2002**, *79*, 213–224. [[CrossRef](#)]
- Yin, Z.; Qin, G.; Guo, L.; Tang, X.; Wang, J.; Li, H. Coupling Antecedent Rainfall for Improving the Performance of Rainfall Thresholds for Suspended Sediment Simulation of Semiarid Catchments. *Sci. Rep.* **2022**, *12*, 4816. [[CrossRef](#)]

30. Wu, Z.Y.; Cheng, D.D.; He, H.; Li, Y.; Zhou, J.H. Research Progress of Composite Drought Index. *Water Resour. Prot.* **2021**, *37*, 36–45.
31. Krajewski, W.F.; Ciach, G.J.; McCollum, J.R.; Bacotiu, C. Initial Validation of the Global Precipitation Climatology Project Monthly Rainfall over the United States. *J. Appl. Meteorol.* **2000**, *39*, 1071–1086. [[CrossRef](#)]
32. Shen, X.; Walker, J.P.; Ye, N.; Wu, X.; Brakhasi, F.; Boopathi, N.; Zhu, L.; Yeo, I.Y.; Kim, E.; Kerr, Y.; et al. Evaluation of the Tau-omega Model over Bare and Wheat-covered Flat and Periodic Soil Surfaces at P-and L-band. *Remote Sens. Environ.* **2022**, *273*, 112960. [[CrossRef](#)]
33. Balti, H.; Abbes, A.B.; Mellouli, N.; Imed, R.F.; Sang, Y.F.; Lamolle, M. A Review of Drought Monitoring with Big Data: Issues, Methods, Challenges and Research Directions. *Ecol. Inform.* **2020**, *60*, 101136. [[CrossRef](#)]
34. Park, S.; Im, J.; Jang, E.; Rhee, J. Drought Assessment and Monitoring Through Blending of Multi-sensor Indices using Machine Learning Approaches for Different Climate Regions. *Agric. For. Meteorol.* **2016**, *216*, 157–169. [[CrossRef](#)]
35. Heydari, H.; Valadan, Z.M.J.; Maghsoudi, Y.; Dehavi, S. An Investigation of Drought Prediction Using Various Remote-sensing Vegetation Indices for Different Time Spans. *Int. J. Remote Sens.* **2018**, *39*, 1871–1889. [[CrossRef](#)]
36. Shen, R.; Huang, A.; Li, B.; Guo, J. Construction of a Drought Monitoring Model Using Deep Learning Based on Multi-source Remote Sensing Data. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *79*, 48–57. [[CrossRef](#)]
37. Cao, J.; Zhang, Z.; Tao, F.; Zhang, L.; Luo, Y.; Zhang, J. Integrating Multi-source Data for Rice Yield Prediction across China Using Machine Learning and Deep Learning Approaches. *Agric. For. Meteorol.* **2021**, *297*, 108275. [[CrossRef](#)]
38. Sardar, V.; Chaudhari, S.; Anchalika, A.; Kakati, A.; Paudel, A.; Bhavana, B.N. Ensemble Learning with CNN and BMO for Drought Prediction. In Proceedings of the 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 7–9 October 2022; pp. 1–6.
39. Kafy, A.A.; Bakshi, A.; Saha, M.; Faisal, A.A.; Almulhim, A.L.; Rahaman, Z.A.; Mohammad, P. Assessment and Prediction of Index Based Agricultural Drought Vulnerability Using Machine Learning Algorithms. *Sci. Total Environ.* **2023**, *867*, 161394. [[CrossRef](#)]
40. Zhao, Y.Y.; Zhang, J.H.; Bai, Y.; Zhang, S.; Yang, S.S.; Henchiri, M.; Seka, A.M.; Nanzad, L. Drought Monitoring and Performance Evaluation based on Machine Learning Fusion of Multi-Source Remote Sensing Drought Factors. *Remote Sens.* **2022**, *14*, 6398. [[CrossRef](#)]
41. Ali, S.; Khorrami, B.; Jehanzaib, M.; Tariq, A.; Ajmal, M.; Arshad, A.; Shafeeque, M.; Dilawar, A.; Basit, I.; Zhang, L. Spatial Downscaling of GRACE Data Based on XGBoost Model for Improved Understanding of Hydrological Droughts in the Indus Basin Irrigation System (IBIS). *Remote Sens.* **2023**, *15*, 873. [[CrossRef](#)]
42. Dikshit, A.; Pradhan, B.; Santosh, M. Artificial Neural Networks in Drought Prediction in the 21st Century—A Scientometric Analysis. *Appl. Soft Comput.* **2022**, *114*, 108080. [[CrossRef](#)]
43. Dikshit, A.; Pradhan, B. Interpretable and Explainable AI (XAI) Model for Spatial Drought Prediction. *Sci. Total Environ.* **2021**, *801*, 149797. [[CrossRef](#)]
44. Ji, Y.H.; Zhou, G.S.; Wang, S.D.; Wang, L.X. Increase in Flood and Drought Disasters during 1500–2000, in Southwest China. *Nat. Hazards* **2015**, *77*, 1853–1861. [[CrossRef](#)]
45. Fu, R.; Chen, R.; Wang, C.; Chen, X.; Gu, H.; Wang, C.; Xu, B.; Liu, G.; Yin, G. Generating High-Resolution and Long-Term SPEI Dataset over Southwest China through Downscaling EEAD Product by Machine Learning. *Remote Sens.* **2022**, *14*, 1662. [[CrossRef](#)]
46. Mei, P.; Liu, J.; Liu, C.; Liu, J.N. A Deep Learning Model and Its Application to Predict the Monthly MCI Drought Index in the Yunnan Province of China. *Atmosphere* **2022**, *13*, 1951. [[CrossRef](#)]
47. Zhang, Z.B.; Yang, Y.; Zhang, X.P.; Chen, Z.J. Wind Speed Changes and Its Influencing Factors in Southwestern China. *Acta Ecol. Sin.* **2014**, *34*, 471–481.
48. Zhang, Q.; Li, Y.Q. Climatic Variation of Rainfall and Rain Day in Southwest China for Last 48 Years. *Plateau Meteorol.* **2014**, *33*, 372–383.
49. Li, Q.; Wang, X.M.; Zhou, G.B.; Zhang, Y.P.; He, Y. Temporal and Spatial Distribution Characteristics of Short-time Heavy Rainfall during Southwest Vortex Rainstorm in Sichuan Basin. *Plateau Meteorol.* **2020**, *39*, 960–972.
50. Zhang, Y.D.; Zhang, X.H.; Liu, S.R. Correlation Analysis on Normalized Difference Vegetation Index (NDVI) of Different Vegetations and Climatic Factors in Southwest China. *Chin. J. Appl. Ecol.* **2011**, *22*, 323–330.
51. Zhao, Q.Q.; Zhang, J.P.; Zhao, T.B.; Li, J.H. Vegetation Changes and Its Response to Climate Change in China Since. *Plateau Meteorol.* **2021**, *40*, 292–301.
52. Gruber, A.; Scanlon, T.; Robin, V.D.S.; Wagner, W.; Dorigo, W. Evolution of the ESA CCI Soil Moisture Climate Data Records and Their Underlying Merging Methodology. *Earth Syst. Sci. Data* **2019**, *11*, 717–739. [[CrossRef](#)]
53. Dorigo, W.; Wagner, W.; Albergel, C.; Albrecht, F.; Balsamo, G.; Brocca, L.; Chung, D.; Ertl, M.; Forkel, M.; Gruber, A.; et al. ESA CCI Soil Moisture for Improved Earth System Understanding: State-of-the Art and Future Directions. *Remote Sens. Environ.* **2017**, *203*, 185–215. [[CrossRef](#)]
54. Guo, X.; Yu, H.B.; Ma, Z.C.; Cao, C.M. Analysis of Spatial and Temporal Variations of Soil Moisture Content and Drought Degree based on MODIS. *Resour. Soil. Water Conserv.* **2019**, *26*, 185–189.
55. Ma, H.X.; Chen, C.C.; Song, Y.Q.; Ye, S.; Hu, Y.M. Analysis of Vegetation Cover Change and Its Driving over the Past Ten Years in Qinghai Province. *Resour. Soil. Water Conserv.* **2018**, *25*, 137–145.
56. GB/T 20481-2017; National Standard of the People’s Republic of China. Meteorological Drought Level. Standardization Administration of China: Beijing, China, 2018.

57. Poornima, S.; Pushpalatha, M. Drought Prediction based on SPI and SPEI with Varying Timescales using LSTM Recurrent Neural Network. *Soft Comput.* **2019**, *23*, 8399–8412. [[CrossRef](#)]
58. Zhao, J.P.; Zhang, X.F.; Liao, C.H.; Bao, H.Y. TVDI based Soil Moisture Retrieval from Remotely Sensed Data Over Large Arid Areas. *Remote Sens. Technol. Appl.* **2011**, *26*, 742–750.
59. Wang, M.C.; Yang, S.T.; Dong, G.T.; Bai, J. Estimating Soil Water in Northern China based on Vegetation Temperature Condition Index. *Arid. Land Geogr.* **2012**, *35*, 446–455.
60. Zhang, M.; Zhang, X.; Hu, G.C.; Wang, N. Applicability Analysis of Remote Sensing based Drought Indices in Drought Monitoring of Apple in Luochuan. *Remote Sens. Technol. Appl.* **2021**, *36*, 187–197.
61. Li, X.H.; Mao, F.Y.; Wang, L.; Yang, J.K. Future Drought Projection of Southwestern China based on CMIP5 Model and MCI Index. In Proceedings of the 2021, 7th International Conference on Hydraulic and Civil Engineering & Smart Water Conservancy and Intelligent Disaster Reduction Forum (ICHCE & SWIDR), Nanjing, China, 6–8 November 2021; pp. 266–276.
62. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
63. Park, S.; Im, J.; Park, S.; Rhee, J. Drought Monitoring using High Resolution Soil Moisture Through Multi-sensor Satellite Data Fusion over the Korean Peninsula. *Agric. For. Meteorol.* **2017**, *237*, 257–269. [[CrossRef](#)]
64. Li, S.; Xu, X. Study on Remote Sensing Monitoring Model of Agricultural Drought based on Random Forest Deviation Correction. *INMATEH-Agric. Eng.* **2021**, *64*, 413–422. [[CrossRef](#)]
65. Gu, Q.Y.; Han, Y.; Xu, Y.P.; Yao, H.Y.; Niu, H.F.; Huang, F. Laboratory Research on Polarized Optical Properties of Saline-alkaline Soil based on Semi-empirical Models and Machine Learning Methods. *Remote Sens.* **2022**, *14*, 226. [[CrossRef](#)]
66. Bharathidasan, S.; Venkataeswaran, C.J. Improving Classification Accuracy Based on Random Forest Model with Uncorrelated High Performing Trees. *Int. J. Comput. Appl.* **2014**, *101*, 26–30. [[CrossRef](#)]
67. Chen, T.Q.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
68. Xiao, Y.; Zhao, W.; Ma, M.; He, K. Gap-free LST Generation for MODIS/Terra LST Product Using a Random Forest-Based Reconstruction Method. *Remote Sens.* **2021**, *13*, 2828. [[CrossRef](#)]
69. Cheng, Y.; Li, Y.X.; Wu, H.P.; Li, F.; Li, Y.Z.; He, L. Reconstructing Modis LST Products Over Tibetan Plateau based on Random Forest. In Proceedings of the IGARSS 2020-2020, IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 6226–6229.
70. Sun, M.; Gong, A.; Zhao, X.; Liu, N.; Si, L.; Zhao, S. Reconstruction of a Monthly 1 km NDVI Time Series Product in China Using Random Forest Methodology. *Remote Sens.* **2023**, *15*, 3353. [[CrossRef](#)]
71. Zhao, W.; Duan, S.B. Reconstruction of Daytime Land Surface Temperatures under Cloud-covered Conditions Using Integrated MODIS/Terra Land Products and MSG Geostationary Satellite Data. *Remote Sens. Environ.* **2020**, *247*, 111931. [[CrossRef](#)]
72. Wang, S.Y.; Zhang, Y.; Meng, X.H.; Song, M.H.; Shang, L.Y.; Su, Y.Q.; Li, Z.G. Fill the Gaps of Eddy Covariance Fluxes Using Machine Learning Algorithms. *Plateau Meteorol.* **2020**, *39*, 1348–1360.
73. Chen, Y.; Ma, L.X.; Yu, D.S.; Feng, K.Y.; Wang, X.; Song, J. Improving Leaf Area Index Retrieval Using Multi-Sensor Images and Stacking Learning in Subtropical Forests of China. *Remote Sens.* **2021**, *14*, 148. [[CrossRef](#)]
74. Zhang, R.; Chen, Z.Y.; Xu, L.J.; Ou, C.Q. Meteorological Drought Forecasting based on a Statistical Model with Machine Learning Techniques in Shaanxi Province, China. *Sci. Total Environ.* **2019**, *665*, 338–346. [[CrossRef](#)] [[PubMed](#)]
75. Li, H.; Zhu, Y. Xgboost Algorithm Optimization Based on Gradient Distribution Harmonized Strategy. *J. Comput. Appl.* **2020**, *40*, 1633.
76. Han, Y.X.; Wu, J.P.; Zhai, B.N.; Pan, Y.X.; Huang, G.M.; Wu, L.F.; Zeng, W.Z. Coupling a Bat Algorithm with Xgboost to Estimate Reference Evapotranspiration in the Arid and Semiarid Regions of China. *Adv. Meteorol.* **2019**, *2019*, 9575782. [[CrossRef](#)]
77. Zhang, B.; Salem, F.K.A.; Hayes, M.J.; Simth, K.H.; Tadesse, T.; Wardlow, B.D. Explainable Machine Learning for the Prediction and Assessment of Complex Drought Impacts. *Sci. Total Environ.* **2023**, *898*, 165509.
78. Babcock, C.; Finely, A.O.; Bradford, J.B.; Kolka, R.K.; Birdsey, R.A.; Ryan, M.G. LiDAR based prediction of forest biomass using hierarchical models with spatially varying coefficients. *Remote Sens. Env.* **2015**, *169*, 113–127. [[CrossRef](#)]
79. Brenning, A. Spatial Cross-validation and Bootstrap for the Assessment of Prediction Rules in Remote Sensing: The R Package Sperrorest. In Proceedings of the 2012, IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 5372–5375.
80. Cracknell, M.J.; Reading, A.M. Geological Mapping Using Remote Sensing Data: A Comparison of Five Machine Learning Algorithms, Their Response to Variations in the Spatial Distribution of Training Data and the Use of Explicit Spatial Information. *Comput. Geosci.* **2014**, *63*, 22–33. [[CrossRef](#)]
81. Sharma, R.C.; Hara, K.; Hirayama, H. A Machine Learning and Cross-Validation Approach for the Discrimination of Vegetation Physiognomic Types Using Satellite Based Multispectral and Multitemporal Data. *Scientifica* **2017**, *2017*, 9806479. [[CrossRef](#)]
82. Ramezan, A.; Warner, C.A.; Maxwell, T.E.A. Evaluation of Sampling and Cross-validation Tuning Strategies for Regional-scale Machine Learning Classification. *Remote Sens.* **2019**, *11*, 185. [[CrossRef](#)]
83. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of Machine-learning Classification in Remote Sensing: An Applied Review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [[CrossRef](#)]

84. Duro, D.C.; Franklin, S.E.; Dubé, M.G. A Comparison of Pixel-based and Object-based Image Analysis with Selected Machine Learning Algorithms for the Classification of Agricultural Landscapes Using SPOT-5 HRG Imagery. *Remote Sens. Environ.* **2012**, *118*, 259–272. [[CrossRef](#)]
85. Stone, M. Cross-validators Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. Ser. B (Methodol.)* **1974**, *36*, 111–133. [[CrossRef](#)]
86. Battineni, G.; Sagaro, G.G.; Nalini, C.; Amenta, F.; Tayebati, S.K. Comparative Machine-learning Approach: A Follow-up Study on Type 2 Diabetes Predictions by Cross-validation Methods. *Machines* **2019**, *7*, 74. [[CrossRef](#)]
87. Shen, R.P.; Guo, J.; Zhang, J.X.; Li, L.X. Construction of a Drought Monitoring Model Using the Random Forest Based on Remote Sensing. *J. Geo-Inf. Sci.* **2017**, *19*, 125–133.
88. Deng, J.H. *Deep Learning—Principles, Models and Practice*; Posts & Telecom Press: Beijing, China, 2021; pp. 47–50.
89. Price, J.C. Using Spatial Context in Satellite Data to Infer Regional Scale Evapotranspiration. *IEEE Trans. Geosci. Remote Sens.* **1990**, *28*, 940–948. [[CrossRef](#)]
90. Wang, R.J.; Li, X.H.; Zhou, R.J.; Wang, L. Applicability Analysis of Three Meteorological Drought Indices in Sichuan Province. *Resour. Environ. Yangtze Basin* **2021**, *30*, 734–744.
91. Jia, H.J. Construction and Application of Remote Sensing Drought Monitoring Model based on Machine Learning in Southwestern China. Master's Thesis, Chengdu University of Information Technology, Chengdu, China, 2022.
92. Xie, W.S.; Zhang, Q.; Li, W.; Wu, B.Y. Analysis of the Applicability of Drought Indexes in the Northeast, Southwest and Middle-lower Reaches of Yangtze River of China. *Plateau Meteorol.* **2021**, *40*, 1136–1146.
93. Duan, H.X.; Wang, S.P.; Feng, J.Y. The National Drought Situation and Its Impact and Causes in 2010. *J. Arid Meteorol.* **2011**, *29*, 126–132.
94. Duan, H.X.; Wang, S.P.; Feng, J.Y. The National Drought Situation and Its Impact and Causes in the Summer of 2011. *J. Arid Meteorol.* **2011**, *29*, 392–400.
95. Duan, H.X.; Wang, S.P.; Feng, J.Y. The National Drought Situation and Its Impact and Causes in 2011. *J. Arid Meteorol.* **2012**, *30*, 136–147.
96. Chen, Y. Remote Sensing Retrieval of Soil Moisture Content Based on Ensemble Learning. Master's Thesis, University of Electronic Science and Technology of China, Chengdu, China, 2020.
97. Jia, Y.Q.; Zhang, B. Spatial-temporal Variability Characteristics of Extreme Drought Events based on Daily SPEI in the Southwest China in Recent 55 Years. *Sci. Geogr. Sin.* **2018**, *38*, 474–483.
98. Wang, C.X.; Zhang, S.Q.; Chen, W.X.; Sun, R. Applicability and Revision of MCI in Sichuan Province. *Chin. Agric. Sci. Bull.* **2019**, *35*, 115–121.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.