



Luting Xu ¹, Yanru Li ², Xiao Wang ¹, Lei Liu ¹, Ming Ma ¹ and Junhui Yang ^{3,*}

- ¹ College of Architecture and Civil Engineering, Chengdu University, Chengdu 610106, China; xuluting@cdu.edu.cn (L.X.); wangxiao@cdu.edu.cn (X.W.); maming@cdu.edu.cn (M.M.)
- ² College of Architecture and Urban-Rural Planning, Sichuan Agricultural University, Chengdu 625014, China; liyanru@sicau.edu.cn
- ³ Chengdu Service Center of Park City Constructure, Chengdu 610084, China
- * Correspondence: yangjunhui@cdu.edu.cn

Abstract: Quantification of shading effects from complex terrain on solar radiation is essential to obtain precise data on incident solar radiation in mountainous areas. In this study, a machine learning (ML) approach is proposed to rapidly estimate the shading effects of complex terrain on solar radiation. Based on two different ML algorithms, namely, Ordinary Least Squares (OLS) and Gradient Boosting Decision Tree (GBDT), this approach uses terrain-related factors as input variables to model and analyze direct and diffuse solar radiation shading rates. In a case study of western Sichuan, the annual direct and diffuse radiation shading rates were most correlated with the average terrain shading angle within the solar azimuth range, with Pearson correlation coefficients of 0.901 and 0.97. The GBDT-based models achieved higher accuracy in predicting direct and diffuse radiation shading rates, with R² values of 0.982 and 0.989, respectively, surpassing the OLS-based models by 0.081 and 0.023. In comparisons between ML models and classic curve-fitting models, the GBDT-based models consistently performed better in predicting both the direct radiation shading rate and the diffuse radiation shading rate, with a standard deviation of residuals of 0.330% and 0.336%. The OLS-based models also showed better performance compared to the curve-fitting models.

Keywords: solar radiation; shading rate; estimation; mountainous area

1. Introduction

With the rapid development of society and the economy, energy issues have become increasingly severe. The development and utilization of renewable energy have emerged as paramount concerns for nations across the world. Solar energy, characterized by its abundance, sustainability, and reliability, has assumed a significant role as a renewable energy source capable of replacing conventional energy sources. Solar energy can be applied in many fields, including architectural design, solar power generation, heat collection system design, and plant growth monitoring.

The utilization of solar energy heavily relies on the availability of precise data on incident radiation. The incident solar radiation is typically influenced by various factors, such as cloud cover, air penetration, and rugged topography [1]. In mountainous regions with complex terrain, the shading caused by the surrounding topography is a key factor affecting the spatial and temporal distribution of solar radiation on the ground. To promote the utilization of solar energy in mountainous areas, it is essential to obtain precise data on incident solar radiation in these areas.

The conventional approach to gathering solar radiation data for a specific region involves the installation of a sufficient number of solar radiation meters in the area. However, in mountainous areas with varied topography, the geographical and economic constraints often severely limit the number of solar radiation monitoring stations. Consequently, relying on ground-based measurements to obtain incident solar radiation data for the entire



Citation: Xu, L.; Li, Y.; Wang, X.; Liu, L.; Ma, M.; Yang, J. A Machine Learning Approach to Estimating Solar Radiation Shading Rates in Mountainous Areas. *Sustainability* 2024, *16*, 931. https://doi.org/ 10.3390/su16020931

Academic Editors: Sameer Al-Dahidi, Mohamed Louzazni and Manoharan Madhiarasan

Received: 24 November 2023 Revised: 17 January 2024 Accepted: 18 January 2024 Published: 22 January 2024



Copyright: © 2024 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/). mountainous region proves unfeasible. An inadequate supply of measured solar radiation data significantly impedes the utilization of solar energy in these areas [2]. Therefore, solar radiation estimation models have become a vital approach to acquire accurate data on incident solar radiation in mountainous regions.

There are many solar radiation estimation models, broadly categorized into three classes based on their calculation principles: analytical models [3–5], empirical models [6–8], and statistical models [9–11]. Analytical models are mathematical models based on physical assumptions, considering the influence of atmospheric components on solar radiation [3]; such models include the work of Threlkeld and Jordan [4] and that of Thevenard and Gueymard [5]. Empirical models establish solar radiation estimation models based on the correlation between meteorological parameters and solar radiation [6]; examples include the models by Iqbal [7] and Bahle et al. [8]. Statistical models iteratively adjust parameter combinations and values between the input variable and the output result to find the optimal parameters based on existing datasets, as in the models by Pang et al. [9], Ozoegwu [10], and Mghouchi et al. [11]. However, these conventional solar radiation estimation models do not consider the shading effects of tall mountains on solar radiation and cannot be directly applied to estimate solar radiation in mountainous areas.

To accurately estimate data on incident solar radiation in mountainous regions, a comprehensive analysis of the shading effects of complex terrain on solar radiation is imperative. The sky-view factor (SVF) is a common parameter used to quantify terrain shading effects on solar radiation [12–16]. SVF is defined as the ratio of the unobstructed surface area to the entire surface area of the hemispherical sky [17]. For mountainous areas with complex terrain, Zhang et al. [18] conducted an in-depth analysis of the terrain shading effects on solar radiation by using SVF, based on high-resolution digital elevation models (DEMs) and atmospheric data products from the Moderate-resolution Imaging Spectroradiometer. This algorithm demonstrated satisfactory performance in validation on the Heihe River Basin, with both the mean bias error percentage (MBE%; -6.2%) and the root mean square difference percentage (RMSD%; 7.5%) below 10% [18]. Vartholomaios [15] developed an efficient machine learning (ML)-based model for urban solar radiation estimation considering terrain shading effects by using the Monte Carlo method to generate 30,000 samples and calculating the SVF and solar radiation data of each sample. This model can achieve near-instant calculation, but the accuracy of its calculation is limited in geometrically complex environments. Although there is a strong correlation between SVF and terrain shading [16,19–23], their relationship is non-linear [24]. This is primarily due to SVF reflecting the percentage of the unobstructed sky hemisphere area relative to the total area, exhibiting a stronger correlation with diffuse radiation shading than with direct radiation shading [15]. Consequently, in heavily shaded areas, estimating solar radiation with SVF would result in a large error.

Using DEM and geographic information system (GIS) platforms to extract terrain data is a common method to quantify mountain shading effects on solar radiation [25–29]. Previously, Dubayah and Rich [30] developed a GIS-based solar radiation estimation model (SolarFlux) after considering the shading effects of mountainous terrain. Because of its simple and convenient operation, this model can be effectively applied in planning, conservation, microclimate, and basic ecology studies. However, this model simplifies the diffuse radiation to isotropic, leading to diminished calculation accuracy of solar radiation. Subsequently, Fu and Rich [31] proposed the concept of the viewshed to quantify the visible and invisible areas of the sky affected by terrain, and they developed the Solar Analyst model. This model assumes diffuse radiation to be anisotropic radiation, leading to more precise calculation results compared to the SolarFlux model. Despite enhancements in model accuracy, the Solar Analyst model is limited to calculating solar radiation within a defined time span. When dealing with small time intervals, the issue of overlapping solar trajectories can arise, resulting in imprecise calculations [32]. In subsequent studies, many scholars have adopted the viewshed concept to develop models for estimating solar

radiation shading levels in diverse environments [28,29], but the limited quantity of raster data continues to pose a challenge in achieving high accuracy.

Although several relevant studies have addressed terrain shading effects on solar radiation, current methods suffer from high computational complexity and limited accuracy, particularly when applied in regions with severe terrain shading. Achieving precise and rapid quantification of terrain shading effects on solar radiation remains a challenging task. Hence, this study proposed an ML-based approach to quickly estimate terrain shading effects on solar radiation solar radiation in mountainous areas. By using this method to rapidly obtain solar radiation shading rates, one can accurately correct the data calculated by conventional models, allowing the precise acquisition of solar radiation data under mountain terrain shading. It could provide essential foundational data for urban site selection and the assessment of rooftop solar potential in mountainous regions.

The study is structured in six sections. Section 1 is the introduction. Section 2 outlines the steps taken to develop the ML model, presenting an analytical approach for assessing terrain shading effects on solar radiation in complex mountainous terrain. Section 3 applies this approach through a case study of western Sichuan, with complex and diverse terrain. The primary results and discussion are presented in Sections 4 and 5 to demonstrate the significance of the research. Finally, Section 6 draws conclusions from the research.

2. Methodology

The methodology for rapidly predicting complex terrain shading effects on solar radiation mainly consists of two parts: the acquisition of training and test sets and the establishment of an ML model. The dataset primarily included terrain factors and solar radiation shading rates. First, based on DEMs, this study used ArcGIS to obtain the terrain factors of the study sites in mountainous areas. Then, the model proposed by Xu et al. [33] was used to predict the solar radiation received by the study sites under terrain shading effects and calculate the solar radiation shading rates. Subsequently, based on the training and test sets, a suitable ML algorithm was chosen to achieve the rapid estimation of direct solar radiation shading rates and diffuse solar radiation shading rates.

2.1. Acquisition of ML Training and Test Sets

2.1.1. Terrain Factors

To accurately describe the spatial relationships between the shading points cast by the surrounding mountains and the study sites, this study introduces the concept of the terrain shading angle, utilizing it as an input parameter of training samples. For a given azimuth around a research point in a mountainous area, the terrain shading angle is defined as the angle between the line connecting the research point to the shading point of the mountain at that azimuth and the horizontal ground, as illustrated in Figure 1. The shading point of the mountain is a point that has shading effects on the solar radiation received by the research point. Based on the geometric relationship between the mountain shading point and the research point, the terrain shading angle (S_i) at a particular azimuth direction of the research point can be calculated by using a trigonometric function. The expression is as follows:

$$\operatorname{Arctan}S_{i} = H_{i}/L_{i} \tag{1}$$

where H_i is the difference in elevation between the research point and the terrain shading point at the specific azimuth angle, while L_i is the horizontal distance between the research point and the terrain shading point.

Based on DEM, with a 1° azimuth interval, this study utilized ESRI ArcGIS 10.2 software's skyline tool to obtain terrain the shading angles for the full 360° around the research point. In this study, it was assumed that all terrain shading angles within a 1° azimuthal interval were identical. Subsequently, average terrain shading angles from the four cardinal directions (east, west, south, and north) and the average terrain shading angle within the annual range of solar azimuth angles were extracted. The summer solstice is the day of the year with the largest range of solar azimuth angles. Therefore, the range of solar azimuth angles throughout the year extends from the azimuth angle of the sun at sunrise on this day to the azimuth angle of the sun at sunset on the same day. The annual range of solar azimuth angles can be calculated using the formula for the solar azimuth angle [34], as shown below:

$$\alpha_{range} = arc\cos(-\frac{\sin\delta}{\cos la}) \tag{2}$$

where δ is the solar declination on the summer solstice and *la* is the local latitude.



Figure 1. Diagram of the terrain shading angle.

2.1.2. Solar Radiation Shading Rates

To further analyze the terrain shading effects on solar radiation, this study introduces the concept of the solar radiation shading rate and uses it as an output parameter of training samples in ML model training. To obtain the solar radiation shading rates, this study uses the model proposed by Xu et al. [33] to calculate the direct solar radiation and diffuse solar radiation received by study sites under terrain shading. Subsequently, the study introduces the concept of the shading rate to quantify the terrain shading effects on direct and diffuse solar radiation. The annual direct solar radiation shading rate is defined as the difference between the annual direct solar radiation received without shading and that with shading, divided by the annual direct radiation received without shading. According to the definition, the expression for the annual direct solar radiation shading rate as proposed in the study is as follows:

$$\eta_{direct} = \frac{E_{direct-0} - E_{direct}}{E_{direct-0}} \times 100\%$$
(3)

where $E_{direct-0}$ is the annual direct solar radiation received without shading and E_{direct} is the annual direct solar radiation received under shading.

Similarly, the concept of diffuse solar radiation shading rate is introduced. The annual diffuse solar radiation shading rate refers to the difference between the annual diffuse solar radiation received without shading and that with shading, divided by the annual diffuse radiation without shading. According to the definition, the study also proposes an expression for the annual diffuse solar radiation shading rate:

$$\eta_{diffuse} = \frac{E_{diffuse-0} - E_{diffuse}}{E_{diffuse-0}} \times 100\%$$
(4)

where $E_{diffuse-0}$ is the annual diffuse solar radiation received without shading and $E_{diffuse}$ is the annual diffuse solar radiation received with shading.

2.2. Selection of ML Algorithm

It has been found that the accuracy of a regression model with multiple independent variables is generally higher than that of a regression model with a single independent variable, and using ML methods to train regression models is an effective solution to multivariate regression problems [35,36]. Therefore, this study sets various terrain factors

as independent variables and selects the direct radiation shading rate and diffuse radiation shading rate as dependent variables. The ML method is used for multivariate regression analysis to deeply investigate the relationships between terrain factors and solar radiation shading rates.

For model training and testing, this study utilized Scikit-learn, a Python-based opensource ML library. Two different algorithms, namely, Ordinary Least Squares (OLS) and Gradient Boosting Decision Tree (GBDT), were chosen for simulation analysis to establish regression models. By comparing the accuracy of the regression models, an appropriate algorithm was selected to develop an ML model that could efficiently quantify the terrain shading effects on solar radiation in mountainous areas.

The Ordinary Least Squares (OLS) algorithm is a linear regression algorithm commonly employed in ML. Its training process is relatively simple, and the resulting model can be expressed directly using a mathematical formula. When the relationship between variables is linear or nearly linear, the prediction results are accurate. However, when the relationship between variables is non-linear, the prediction results might not be accurate. Due to the simple calculation logic of the OLS model, its prediction accuracy is often lower than that of models fitted by more complex regression algorithms, such as ensemble algorithms.

To ensure the prediction accuracy of the model, the study also uses an ensemble algorithm, Gradient Boosting Decision Tree (GBDT). GBDT combines multiple regression decision tree estimators through gradient boosting, increasing the robustness and accuracy of the overall model. It is considered to be one of the best-performing ML methods [37].

3. Case Implementation

In this study, the Western Sichuan Plateau was chosen as a representative case to apply and evaluate the method proposed above. Situated in the transition area between the Qinghai-Tibet Plateau and the Sichuan Basin, the Western Sichuan Plateau is characterized by high altitude and abundant solar energy resources. However, the terrain there is complex and varied. In western Sichuan, the solar radiation received by the towns scattered among the mountains is significantly affected by the shading effects of the surrounding terrain (Figure 2).



Figure 2. Real image of a town on the Western Sichuan Plateau.

3.1. Terrain Factors of Western Sichuan

Based on the distribution of towns on the Western Sichuan Plateau, this study selected 139 town study sites and extracted terrain factors as inputs for training samples. First, based on DEM, ArcGIS was used to extract the terrain shading angles around the study sites at 1° interval. Taking Jianshe Town in western Sichuan as an example, the terrain shading angles around the research point are shown in Figure 3. In the diagram, the angular coordinate represents the position of the mountains relative to the town (-180° to 180°), where 0° signifies that the mountain is to the south of the town, while $\pm 180^{\circ}$ indicates that the mountain is to the town. The radial axis represents the slope of the

mountain (0° to 40°), and the length of the line segment denotes the terrain shading angle at that direction. A longer line segment implies a greater the terrain shading angle at the azimuth angle. Subsequently, based on the terrain shading angles around the 139 town study sites, the average terrain shading angle within the range of solar azimuth angles and the terrain shading angles in the four different directions are calculated.





3.2. Solar Radiation Shading Rates

In rugged mountain areas, varying mountain shapes result in different degrees of shading effects on solar radiation. According to the topographical characteristics of western Sichuan, the direct radiation shading rates and diffuse radiation shading rates of the study sites in 139 towns were calculated by using the solar radiation model of Xu et al. [33]. These shading rates were considered output variables of training samples, as depicted in Figure 4. The average annual direct radiation shading rate of the 139 towns in western Sichuan was 4.6%, and the average annual diffuse radiation shading rate was 8.3%. Among these, the direct radiation shading rates of 57 urban study sites exceeded 5%, and the diffuse radiation shading rates of 95 urban study sites exceed 5%.



Figure 4. Solar radiation shadings rate of 139 towns in western Sichuan.

3.3. Selection of Input-Parameter Combinations

To determine the optimal combination of input variables for ML, the correlations between various terrain factors and solar radiation shading rates were analyzed. Five terrain factors of the 139 towns were used as input variables, with the annual direct radiation shading rates and diffuse radiation shading rates as output variables. The correlations between the five terrain factors and the annual direct radiation shading rates and diffuse radiation shading rates were analyzed.

Figure 5 illustrates the distribution of solar radiation shading rates in western Sichuan under the influence of the five terrain factors and the Pearson correlation coefficients (P) between shading rates and terrain factors. This figure demonstrates that there were linear relationships between the five terrain factors and the annual solar radiation shading rates. Obviously, the annual direct radiation shading rate and the diffuse radiation shading rate exhibited the strongest correlation with the average terrain shading angle within the solar azimuth range, with Pearson correlation coefficients of 0.901 and 0.971, respectively. The annual direct radiation shading rate and the diffuse solar radiation shading rate had the lowest correlation with the average shading angle in the northerly direction, with Pearson correlation coefficients of 0.517 and 0.786, respectively.



Figure 5. Correlation analysis of terrain shading factors and solar radiation shading rates.

According to the results of the correlation analysis, selecting appropriate input variables for model training could enhance the model calculation accuracy. During the model training process, an excessive number of input variables may lead to model overfitting, compromising simulation accuracy. Therefore, to mitigate overfitting, the number of input variables was increased in turn based on the degree of correlation between the five terrain factors and solar radiation shading rates during model training. The input variables were added in the following order: the average terrain shading angle within the solar azimuth range, the terrain shading angle in the east, the terrain shading angle in the west, and the terrain shading angles in the south and north. Throughout the training process, the input variables were divided into five combinations, as shown in Table 1.

Table 1. Different combination groups of input variables.

	S _{solar}	S _{east}	Swest	S_{south}	S _{north}
Group 1	\checkmark				
Group 2	\checkmark	\checkmark			
Group 3	\checkmark	\checkmark	\checkmark		
Group 4	\checkmark	\checkmark	\checkmark	\checkmark	
Group 5		\checkmark		\checkmark	

3.4. GBDT Algorithm Hyperparameter Optimization

When using the GBDT algorithm for model training, it is crucial to adjust and optimize the hyperparameters, including the n-estimators and learning rate. Among them, the nestimators represent the number of iterations for decision tree models. Generally, if the parameter is too small, it may lead to underfitting, and if it is too large, it might result in overfitting. Thus, selecting a moderate value is imperative, taking the learning rate into consideration. In this study, after multiple adjustments and careful consideration of the simulation accuracy, 200 was selected as the n-estimator. The learning rate is the weightshrinking coefficient for each decision tree model, ranging from 0 to 1. Although a smaller learning rate can reduce overfitting, it requires more iterations of the decision tree model for overall training. After multiple parameter adjustment tests, 0.1 was chosen as the learning rate for training in this study. The "subsample" refers to the proportion of subsampling during training. A small subsample value can prevent overfitting but may increase the sample bias. In this study, subsampling was not used, and the subsample value was set to 1. The parameter Loss was used to select the loss function, encompassing mean squared error ("ls"), absolute loss ("lad"), Huber loss ("huber"), and quantile loss ("quantile"). Due to the small sample size in this study, the "ls" parameter was used to improve the training.

Apart from the hyperparameters of the boosting frameworks, several hyperparameters of each individual decision tree estimator also need to be adjusted to establish the GBDT model. The "max depth" represents the maximum depth of the decision tree. After testing, a value of 2 was chosen as the max depth to optimize the model accuracy. The parameter "min-sample-split" denotes the minimum number of samples required for internal node splitting. If the number of the node samples is less than this parameter, no attempt is made to split the node further. Owing to the limited number of samples in this study, this parameter was set to 2. The parameter "min-sample-leaf" signifies the minimum number of samples for a leaf node. If the number of leaf nodes is smaller than the number of samples, it may be pruned along with the sibling nodes. To improve the precision of the model, the value was set to 1.

4. Results

4.1. OLS Algorithm

Based on various terrain shading factors, the OLS algorithm was used to fit the annual direct solar radiation shading rates and the annual diffuse solar radiation shading rates. To verify the model accuracy, the 139 samples were divided into training and test sets in a 5:1 ratio. The best model was selected through comparative analysis of computational errors of models obtained using five different input-variable combinations.

Figure 6 displays comparisons between the simulated values obtained from the OLS algorithm using five different input-variable combinations and the reference values. For the annual direct radiation shading rate, the model's accuracy was highest when utilizing the variables of group 4 as inputs, with an R² of 0.900. This group includes the average terrain shading angle within the solar azimuth range and the terrain shading angles to the east, west, and south. The optimal formula for calculating the annual direct solar radiation shading rate (R_{direct}) as obtained by the OLS algorithm is as follows:

$$R_{direct} = 0.014433S_{solar} - 0.00186S_{east} - 0.00217S_{west} - S_{south}$$
(5)

where S_{solar} is the average terrain shading angle within the solar azimuth range, while S_{east} , S_{west} , and S_{south} are the shading angles to the east, west, and south, respectively.

For the annual diffuse solar radiation shading rate, when using the variables of group 5 as inputs, the model's accuracy was highest with an R^2 of 0.967. This group includes the average terrain shading angle within the solar azimuth range and the terrain shading angles in the four different directions. Therefore, the optimal formula for calculating the annual diffuse solar radiation shading rate ($R_{diffuse}$) as obtained by the OLS algorithm is as follows:

$R_{diffuse} = 0.006468S_{solar} + 0.00087S_{east} + 0.000165S_{west} - 0.00076S_{south} + 0.001675S_{north}$ (6)



where S_{north} represents the shading angle in the northerly direction.

Figure 6. Simulation values of the OLS-based models trained by using 5 groups of input variables vs. reference values.

4.2. GBDT Algorithm

Based on various terrain shading factors, the GBDT algorithm was used to fit the annual direct solar radiation shading rate and the annual diffuse solar radiation shading rate. As with the OLS algorithm, to validate the accuracy of the GBDT-based models, the 139 samples were also divided into training and test sets in a 5:1 ratio.

Figure 7 demonstrates comparisons between simulated values obtained from GBDTbased models trained with the five groups of input variables and the reference values. For the annual direct solar radiation shading rate, the model achieved the highest accuracy using group 4 variables as inputs; these variables included the average terrain shading angle within the solar azimuth range and the shading angles to the east, west, and south, with an R^2 of 0.982. Similarly, for the annual diffuse solar radiation shading rate, the model attained the highest accuracy using group 5 variables, encompassing the average terrain shading angle within the solar azimuth range and the shading angles in all four cardinal directions, with an R^2 of 0.989.



Figure 7. The GBDT-based models' simulation values calculated by using 5 groups of input variables vs. reference values.

5. Discussion

In order to further investigate the intrinsic relationship between the solar radiation shading rates and the terrain shading factor, this study conducted a comparison between ML models and traditional curve-fitting models. Based on the analysis results in the form of Pearson correlation coefficients, it was evident that the direct and diffuse radiation shading rates were most closely related to S_{solar} . Thus, this study set S_{solar} as the independent variable to perform univariate curve fitting. During the fitting process, eight different curve equations were selected, and error statistics were calculated, with the analysis results shown in Table 2. From the table, it is evident that for the estimation of the annual direct and diffuse radiation shading rates, the power-function equations performed the best, with R^2 values of 0.887 and 0.961, respectively. The power-function equations of the annual direct solar radiation shading rate (R_{direct}) and the annual diffuse solar radiation shading rate (R_{direct}) are as follows:

$$R_{direct} = 0.00001332 S_{solar}^{2.894} \tag{7}$$

$$R_{diffuse} = 0.000524 S_{solar}^{1.84} \tag{8}$$

Table 2.	Comparison	of the estimation	tion errors	for solar	radiation	shading	rates l	based or	different
curve-fit	ting models.								

F (1	Direct Shading	Diffuse Shading		
Functions	R ²	R ²		
Logarithmic curve	0.562	0.767		
Inverse function	0.281	0.46		
Quadratic curve	0.829	0.944		
Cubic curve	0.829	0.946		
Composite curve	0.763	0.863		
Power function	0.887	0.961		
S-curve	0.792	0.826		
Logistic	0.765	0.862		

After that analysis, a residual comparison analysis was conducted. By comparing the residuals of the solar radiation shading rate models trained by the curve-fitting, OLS, and GBDT algorithms, a comprehensive analysis was conducted to assess the accuracy of the models. The residual comparison of the direct radiation shading rate simulated by models using these three algorithms is illustrated in Figure 8. For the direct solar radiation shading rate, the R² of the GBDT-based model exceeded that of the OLS-based model by 0.081 and exceeded that of the curve-fitting model by 0.094. The standard deviations of the residuals (SD) for the curve-fitting model, OLS model, and GBDT model were 1.317%, 1.260%, and 0.330%. The residual comparison indicated that the predictions of the GBDTbased model were more stable than those of the other two models. The residual values of the GBDT model consistently remained no greater than 1%, with only a few above 1%. The predicted results of the OLS-based models were influenced by the reference values. When the reference value exceeded 12%, the absolute value of the residual might also exceed 2%. For the curve-fitting model, the error was relatively even across the entire range of reference values, and the number of residual values greater than 2% was higher with this model than with the other two types of ML models.



Figure 8. Residual difference comparison of the simulated values from the direct radiation shading rate prediction models.

Figure 9 displays a residual comparison of the diffuse radiation shading rates simulated by models using the curve-fitting, OLS, and GBDT algorithms. For the diffuse radiation shading rates, the GBDT-based model outperformed the OLS-based model and the curve-fitting model, presenting a higher R^2 value by 0.023 and 0.028, respectively. The difference in accuracy among the three algorithms for predicting the diffuse radiation shading rate was smaller than that for predicting the direct solar radiation shading rate. The standard deviations of the residuals for the curve-fitting model, OLS model, and GBDT model were 1.313%, 1.000%, and 0.336%. The SD of the GBDT-based model was still below 0.5%, and the SDs of the OLS and curve fitting models were larger than 1%. The residual values of the predicted diffuse solar radiation shading rates using the GBDT-based model consistently remained no greater than 1%, with only a few above 1%. The predicted diffuse radiation shading rate of the OLS-based models was also affected by the reference values. When the reference value was greater than 15% or close to 0%, the absolute value of the residuals could be greater than 2%. For the curve-fitting model, the error was similarly uniform across the entire range of reference values, and the quantity of residual values exceeding 2% for this model was more numerous than for the other two types of ML models.



Figure 9. Residual difference comparison of the simulated values from the diffuse radiation shading rate prediction models.

From the aforementioned error comparisons, it can be seen that the direct and diffuse radiation shading rate models based on the GBDT algorithm perform best. The computational precision of the direct and diffuse radiation shading rate models based on the OLS algorithm is only slightly higher than that of the curve fitting equation. Thus, in practical applications, an appropriate algorithm for estimating solar radiation shading rates can be selected according to the actual precision requirement.

Although using the SVF concept to analyze the shading effects on solar radiation is a commonly used method, it only contains information regarding the shading percentage, without giving the precise shading location details [17]. In instances of severe shading, the calculation of solar radiation shading using SVF would cause great errors [15]. Based on the calculation principle of radiance [38] or the concept of the viewshed [31], shading effects can also be quantitatively analyzed. In order to estimate the solar radiation data of the research point under shading, Rutten [39] developed a Grasshopper plugin in Rhinoceros3D 7.5 software based on the principle of radiance calculation. Subsequently, Roudsari and Pak [40] further developed the related complementary plugins for Grasshopper. When calculating the shading effects on solar radiation by the Grasshopper plugin, the sky hemisphere was divided into 145 parts [41]. Zhang et al. [42] developed a model for calculating instantaneous solar radiation under mountainous terrain shading based on the viewshed by dividing the viewshed grid resolution into 16 azimuth and altitude angle parts. However, the finite number of sky hemisphere divisions restricts the calculation accuracy.

The solar radiation shading rate of the training set of this study is calculated based on the model of Xu et al. [33]. It divides the viewshed grid resolution into 360 parts in the azimuth and 90 parts in the altitude angle. The calculation of diffuse solar radiation data under shading effects takes place through double integration on the unblocked surface of the sky dome [33]. Compared with the typical measured annual meteorological data, the R² of the hourly solar radiation calculation errors of the model of Xu et al. [33] exceeds 0.97. Therefore, the method presented in this study is not only precise in the ML process but also has accurate training and test sets. In summary, the solar radiation shading rate estimation model proposed by the research not only has a fast calculation speed but also has satisfactory calculation accuracy compared with the existing shading quantification analysis models.

However, there are also some limitations to this study. Specifically, the focus was only on the annual solar radiation shading rates, without considering variations across different seasons. In addition, the solar radiation shading rates of surfaces with varying orientations have not been involved. In future work, the solar radiation shading rates of surfaces with varying orientations during different time periods need to be further studied.

6. Conclusions

Based on mountainous terrain factors, this study proposed an ML method for solar radiation shading rate prediction in mountainous areas with complex terrain by using the OLS and GBDT algorithms. Error analysis was conducted to select the optimal models, enabling accurate and rapid predictions of both the direct radiation shading rate and the diffuse radiation shading rate.

The Western Sichuan Plateau was chosen as a representative case to establish a rapid estimation model of solar radiation shading rates. The complex terrain in western Sichuan has an obvious shading impact on solar radiation, and the average direct radiation shading rate of the 139 towns in western Sichuan is 4.6%, while the average diffuse radiation shading rate is 8.3%. According to the correlation analysis between various terrain factors and solar radiation shading rates, it was revealed that the annual direct and diffuse radiation shading angle within the solar azimuth range, with Pearson correlation coefficients of 0.901 and 0.971, respectively. The annual direct radiation shading rate and the diffuse solar radiation shading rate have the lowest correlation with the average shading angle in the north direction, with Pearson coefficients of 0.517 and 0.786, respectively.

During the model development process, a comparative analysis was performed using five sets of input variables. For western Sichuan, the R^2 values of the optimal OLSbased models for direct solar radiation shading rate and diffuse radiation shading rate are 0.900 and 0.967, respectively. However, the R² values for the optimal GBDT-based models are higher, at 0.982 and 0.989, respectively. Furthermore, the study made a comparison between the classic curve-fitting model and the ML model. Since the direct and diffuse radiation shading rates are most closely related to S_{solar} , this study set S_{solar} as the independent variable and conducted a single-factor curve fitting. The resulting equations for the annual shading rates of direct and diffuse radiation yielded R^2 values of 0.887 and 0.961, respectively. It is found that for the direct solar radiation shading rate prediction, the standard deviations of residuals for the curve-fitting model, OLS model, and GBDT model are 1.317%, 1.260%, and 0.330%. And for the diffuse solar radiation shading rate prediction, the standard deviations of residuals for the curve-fitting model, OLS model, and GBDT model are 1.313%, 1.000%, and 0.336%. Therefore, no matter whether it is used for the direct radiation shading rate prediction or the diffuse radiation shading rate prediction, the GBDTbased models always perform best, and the OLS-based models perform slightly better than the curve-fitting models. In practical applications, an appropriate algorithm for estimating solar radiation shading rates can be selected according to the actual precision requirements.

In summary, the solar radiation shading rate estimation model proposed by the research not only has a fast calculation speed but also has satisfactory calculation accuracy compared with the existing shading quantification analysis models. Based on the solar radiation shading rates predicted by this method, solar radiation data obtained by conventional models that do not consider terrain shading effects can be corrected to obtain precise solar radiation data in mountainous areas. This study could effectively address the issue of the insufficient solar radiation data in mountainous areas, thereby providing crucial fundamental data for the field of urban site selection and rooftop solar energy utilization.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su16020931/s1.

Author Contributions: Conceptualization, L.X.; methodology, L.X.; software, L.X.; validation, L.X.; formal analysis, Y.L. and L.L.; investigation, L.L.; resources, X.W.; writing—original draft preparation, L.X.; writing—review and editing, L.X., Y.L. and M.M.; supervision, J.Y.; funding acquisition, L.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 52208006.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the supplementary material named "Data".

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- 1. Singh, S.K.; Lohani, B.; Arora, L.; Choudhary, D.; Nagarajan, B. A visual-inertial system to determine accurate solar insolation and optimal PV panel orientation at a point and over an area. *Renew. Energy* **2020**, *154*, 223–238. [CrossRef]
- Aguilar, C.; Herrero, J.; Polo, M.J. Topographic effects on solar radiation distribution in mountainous watersheds and their influence on reference evapotranspiration estimates at watershed scale. *Hydrol. Earth Syst. Sci.* 2010, 14, 2479–2494. [CrossRef]
- 3. Sun, X.; Bright, J.M.; Gueymard, C.A.; Acord, B.; Wang, P.; Engerer, N.A. Worldwide performance assessment of 75 global clear-sky irradiance models using Principal Component Analysis. *Renew. Sustain. Energy Rev.* **2019**, *111*, 550–570. [CrossRef]
- 4. Threlkeld, J.L.; Jordan, R.C. Direct Solar Radiation Available on Clear Days. Heat Pip. Air Cond. 1957, 29, 135–145.
- Thevenard, D.; Gueymard, C.A. Updating the ASHRAE Climatic Data for Design and Standards; American Society of Heating, Refrigerating and Air-Conditioning Engineers: Atlanta, GA, USA, 2009.

- 6. Zhou, Y. Research on the Development of the Daily Solar Radiation Estimation Models and Outdoor Design Radiation. Ph.D. Thesis, Xi'an University of Architecture and Technology, Xi'an, China, 2019.
- Iqbal, M. Correlation of average diffuse and beam radiation with hours of bright sunshine. Sol. Energy 1979, 19, 169–173. [CrossRef]
- 8. Bahel, V.; Bakhsh, H.R.S. A correlation for estimation of global solar radiation. *Energy* **1987**, *12*, 131–135. [CrossRef]
- Pang, Z.; Niu, F.; O'Neill, Z. Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. *Renew. Energy* 2020, 156, 279–289. [CrossRef]
- 10. Ozoegwu, C.G. Artificial neural network forecast of monthly mean daily global solar radiation of selected locations based on time series and month number. *J. Clean. Prod.* **2019**, *216*, 1–13. [CrossRef]
- 11. El Mghouchi, Y.; Chham, E.; Zemmouri, E.M.; El Bouardi, A. Assessment of different combinations of meteorological parameters for predicting daily global solar radiation using artificial neural networks. *Build. Environ.* **2019**, 149, 607–622. [CrossRef]
- 12. Redweik, P.; Catita, C.; Brito, M. Solar energy potential on roofs and facades in an urban landscape. *Sol. Energy* **2013**, *97*, 332–341. [CrossRef]
- Catita, C.; Redweik, P.; Pereira, J.; Brito, M.C. Extending solar potential analysis in buildings to vertical facades. *Comput. Geosci.* 2014, 66, 1–12. [CrossRef]
- 14. Brito, M.C.; Freitas, S.; Guimares, S.; Catita, C.; Redweik, P. The importance of facades for the solar PV potential of a Mediterranean city using LiDAR data. *Renew. Energy* 2017, 111, 85–94. [CrossRef]
- 15. Vartholomaios, A. A machine learning approach to modelling solar irradiation of urban and terrain 3D models. *Comput. Environ. Urban Syst.* **2019**, *78*, 101387. [CrossRef]
- 16. Robinson, D. Urban morphology and indicators of radiation availability. Sol. Energy 2006, 80, 1643–1648. [CrossRef]
- 17. Johnson, G.; Watson, I. The determination of view-factors in urban canyons. J. Clim. Appl. Meteorol. 1984, 23, 329–335. [CrossRef]
- Zhang, Y.; Li, X.; Bai, Y. An integrated approach to estimate shortwave solar radiation on clear-sky days in rugged terrain using MODIS atmospheric products. *Sol. Energy* 2015, *113*, 347–357. [CrossRef]
- 19. Kämpf, J.H.; Robinson, D. Optimisation of building form for solar energy utilisation using constrained evolutionary algorithms. *Energy Build.* **2010**, *42*, 807–814. [CrossRef]
- 20. Ketterer, C.; Matzarakis, A. Mapping the physiologically equivalent temperature in urban areas using artificial neural network. *Landsc. Urban Plan.* **2016**, *150*, 1–9. [CrossRef]
- 21. Chatzidimitriou, A.; Yannas, S. Street canyon design and improvement potential for urban open spaces; the influence of canyon aspect ratio and orientation on microclimate and outdoor comfort. *Sustain. Cities Soc.* **2017**, *33*, 85–101. [CrossRef]
- Poon, K.H.; Kämpf, J.H.; Tay, S.E.R.; Wong, N.H.; Reindl, T.G. Parametric study of URBAN morphology on building solar energy potential in Singapore context. Urban Clim. 2020, 33, 100624. [CrossRef]
- Lan, H.; Gou, Z.; Xie, X. A simplified evaluation method of rooftop solar energy potential based on image semantic segmentation of urban streetscapes. Sol. Energy 2021, 230, 912–924. [CrossRef]
- Mohajeri, N.; Gudmundsson, A.; Kunckler, T.; Upadhyay, G.; Assouline, D.; Kämpf, J.H.; Scartezzini, J.L. A solar-based sustainable urban design: The effects of city-scale street-canyon geometry on solar access in Geneva, Switzerland. *Appl. Energy* 2019, 240, 173–190. [CrossRef]
- 25. Hetrick, W.A.; Rich, P.M.; Barnes, F.J.; Weiss, S.B. GIS-based solar radiation flux models. *Am. Soc. Photogramm. Remote Sens. Tech. Pap.* **1993**, *3*, 132–143.
- 26. Rich, P.M. Characterizing plant canopies with hemispherical photographs. Remote Sens. Rev. 1990, 5, 13–29. [CrossRef]
- 27. Hofierka, J.; Suri, M. The solar radiation model for Open source GIS: Implementation and applications. In Proceedings of the Open Source GIS—GRASS Users Conference, Trento, Italy, 11–13 September 2002.
- Ivanova, S.M.; Gueymard, C.A. Simulation and applications of cumulative anisotropic sky radiance patterns. Sol. Energy 2019, 178, 278–294. [CrossRef]
- Liao, W.; Heo, Y.; Xu, S. Simplified vector-based model tailored for urban-scale prediction of solar irradiance. Sol. Energy 2019, 183, 566–586. [CrossRef]
- 30. Dubayah, R.; Rich, P.M. Topographic solar-Radiation models for GIS. Int. J. Geogr. Inf. Syst. 1995, 9, 405–413. [CrossRef]
- 31. Fu, P.; Rich, P.M. Design and implementation of the Solar Analyst: An ArcView extension for modeling solar radiation at landscape scales. In Proceedings of the 19th Annual ESRI User Conference, San Diego, CA, USA, 26–30 July 1999.
- ESRI. Solar Radiation Tools. ArcGis 10.2 Help. 2013. Available online: https://resources.arcgis.com/en/help/main/10.2/index. html#//009z00000000000000000 (accessed on 4 October 2014).
- Xu, L.; Long, E.; Wei, J.; Cheng, Z.; Zheng, H. A new approach to determine the optimum tilt angle and orientation of solar collectors in mountainous areas with high altitude. *Energy* 2021, 237, 121507. [CrossRef]
- 34. ASHRAE. ASHRAE's Handbook of Fundamentals; American Society of Heating, Refrigerating and Air-Conditioning Engineers: Atlanta, GA, USA, 2017; Chapter 14.
- Aiken, L.S.; West, S.G.; Pitts, S.C. Multiple Linear Regression. In *Handbook of Psychology*; John Wiley & Sons: Hoboken, NJ, USA, 2003; pp. 481–507. [CrossRef]
- Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* 2015, 71, 804–818. [CrossRef]

- 37. Wang, J.; Li, P.; Ran, R.; Che, Y.; Zhou, Y. A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree. *Appl. Sci.* 2018, *8*, 689. [CrossRef]
- Ward, G.J. The RADIANCE lighting simulation and rendering system. In Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH, Orlando, FL, USA, 24–29 July 1994; pp. 459–472.
- Rutten, D. Grasshopper. Generative Modeling for Rhino. 2012. Available online: https://www.grasshopper3d.com/ (accessed on 23 November 2023).
- Sadeghipour Roudsari, M.; Pak, M. Ladybug: A parametric environmental plugin for grasshopper to help designers create an environmentally-conscious design. In Proceedings of the BS 2013: 13th Conference of the International Building Performance Simulation Association, Chambery, France, 25–28 August 2013; pp. 3128–3135.
- 41. Robinson, D.; Stone, A. Irradiation modelling made simple: The cumulative sky approach and its applications. In Proceedings of the Conference on Passive and Low Energy Architecture, Eindhoven, The Netherlands, 19–22 September 2004; pp. 19–22.
- 42. Zhang, S.; Li, X.; She, J.; Peng, X. Assimilating remote sensing data into GIS-based all sky solar radiation modeling for mountain terrain. *Remote Sens. Environ.* **2019**, 231, 111239. [CrossRef]

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.