

Article

A Fine-Grained Simulation Study on the Incidence Rate of Dysentery in Chongqing, China

Jian Hao ^{1,2} and Jingwei Shen ^{1,2,*} 

¹ Chongqing Jinpo Mountain Karst Ecosystem National Observation and Research Station, School of Geographical Sciences, Southwest University, Chongqing 400715, China; hj1358168939@email.swu.edu.cn

² Chongqing Engineering Research Center for Remote Sensing Big Data Application, School of Geographical Sciences, Southwest University, Chongqing 400715, China

* Correspondence: sjwgis@swu.edu.cn

Abstract: Dysentery is still a serious global public health problem. In Chongqing, China, there were 37,140 reported cases of dysentery from 2015 to 2021. However, previous research has relied on statistical data of dysentery incidence rate data based on administrative regions, while grained scale products are lacking. Thus, an initialized gradient-boosted decision trees (IGBDT) hybrid machine learning model was constructed to fill this gap in grained scale products. Socioeconomic factors, meteorological factors, topographic factors, and air quality factors were used as inputs of the IGBDT to map the statistical dysentery incidence rate data of Chongqing, China, from 2015 to 2021 on the grid scale. Then, dysentery incidence rate grained scale products (1 km) were generated. The products were evaluated using the total incidence of Chongqing and its districts, with resulting R^2 values of 0.7369 and 0.5439, indicating the suitable prediction performance of the model. The importance and correlation of factors related to the dysentery incidence rate were investigated. The results showed that socioeconomic factors had the main impact (43.32%) on the dysentery incidence rate, followed by meteorological factors (33.47%). The Nighttime light, normalized difference vegetation index, and maximum temperature showed negative correlations, while the population, minimum and mean temperature, precipitation, and relative humidity showed positive correlations. The impacts of topographic factors and air quality factors were relatively weak.

Keywords: dysentery; initialized gradient-boosted regression trees; downscaling; fine-grained scale product



Citation: Hao, J.; Shen, J. A

Fine-Grained Simulation Study on the Incidence Rate of Dysentery in Chongqing, China. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 459. <https://doi.org/10.3390/ijgi12110459>

Academic Editors: Christos Chalkias, Demosthenes Panagiotakos and Wolfgang Kainz

Received: 26 September 2023

Revised: 3 November 2023

Accepted: 7 November 2023

Published: 9 November 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

Dysentery is a water- and food-borne infectious disease [1]. It is an intestinal infectious disease transmitted through contaminated water, food, or human contact. It includes bacillary dysentery and amebic dysentery [1,2]. Dysentery is also a global public health problem with high contagiousness and complex transmission, affecting people of all ages, especially in developing countries [3–5]. According to a report from the World Health Organization (<https://www.who.int>, accessed on 2 November 2023), diarrhoeal is the second leading cause of death among children under the age of five. Approximately 525,000 children under the age of five die from diarrhoeal each year, and there are approximately 1.7 billion cases of diarrhoeal in children worldwide every year. And, dysentery is a significant subtype of diarrhea. In China, dysentery is still classified as a Class A or B legally notifiable infectious disease. According to the report on legally notifiable diseases by the National Health Commission of China in 2021 (<http://www.nhc.gov.cn/jkj/>, accessed on 8 July 2023), the number of reported cases of dysentery was 50,403, with a national incidence rate of 3.5752 (1/10⁵, per 100,000 population).

The United Nations introduced the Sustainable Development Goals in 2015, and within the “Good Health and Well-being” goal, there is a clear aim to eliminate epidemic

diseases by 2030. Given that dysentery remains a prevalent infectious disease worldwide, it is crucial to study and analyze the spatial and temporal clustering areas, development trends, and related factors of dysentery. Although there are many studies related to dysentery, such as spatial and temporal distribution studies [5–9], prediction of dysentery incidence [10], and related influencing factors [3,11–19], these studies were based on the incidence rate of dysentery statistics data within administrative regions. Although statistical dysentery incidence rate data can intuitively represent geographical phenomena and spatial distribution and be convenient for spatial relationships and statistical analysis, it is limited by the use of predefined statistical units [20]. As such, it cannot express the situation within units in a fine-grained manner and cannot achieve more detailed research. Grid data can provide this required detail [21,22], and it has the advantage of integrating various related factors, such as topographic and meteorological factors. Therefore, there is a demand for mapping dysentery incidence rate data to obtain a fine-grained distribution of dysentery incidence rates.

The factors related to dysentery incidence should be investigated first to obtain the fine-grained distribution of dysentery incidence rates. As for the factors affecting dysentery, they can be broadly summarized as socioeconomic factors [1–3,15,23,24], such as population and economic development status, meteorological factors [13,14,16–19,24], such as precipitation and temperature, topographical factors [11,15], such as slope and elevation differences, and air quality factors [25,26], such as PM_{2.5} and PM₁₀. Therefore, this study believes that using these factors for research is appropriate. However, these factors have strong nonlinear relationships with the dysentery incidence rate, so machine learning methods are a suitable approach for mapping the dysentery incidence rate to the grid scale.

To date, some scholars have used machine learning methods to map statistical data to the grid scale, such as using a random forest (RF) model to map population census data to a 1 km grid scale [21], using an ensemble approach for fine-scale dynamic population mapping [27] using a convolutional neural network method to map gross domestic product (GDP) statistics data to a grid scale [22], and using an RF model to map NO₂ concentration data to a grid scale [28]. Compared to traditional multiple linear regression methods, machine learning algorithms such as RF, gradient-boosted decision trees (GBDT), and neural networks are better at explaining the nonlinear relationships between variables. The initialized gradient-boosted decision trees (IGBDT) [29] is a hybrid machine learning model that combines the RF [30] and GBDT [31,32], thereby integrating the strengths of both algorithms. This model has suitable interpretability, allowing for an explanation of the impact of each feature on the target variable. These advantages make IGBDT highly generalizable and capable of explaining nonlinear relationships, thus allowing for more accurate predictions of the dysentery incidence rate.

The main contributions of this paper lie in two aspects. An IGBDT hybrid machine learning model is used to downscale the annual dysentery incidence rate data to a 1 km grid scale. Second, the importance levels of and correlations among various factors affecting the incidence rate of dysentery are revealed.

The remainder of this article is organized as follows. In the Section 2, the data sources, data processing, and integration process used herein are introduced. The Section 3 describes the workflow of this study and the IGBDT model. In the Section 4, the evaluation results of the constructed IGBDT model are presented. The dysentery incidence rate grained scale products (1 km) of Chongqing generated by the model are described, and the importance levels of and correlations of various relevant factors are discussed. In the fifth part, the IGBDT model is compared with other models, and the importance levels of and correlations among the influencing factors obtained in this study are compared with other studies. Finally, the contributions, limitations, and future prospects of this study are summarized.

2. Materials and Methods

2.1. Study Area

The study area considered in this work is Chongqing, a municipality located in south-western China and one of the four municipalities directly under the central government of China. Chongqing has 38 districts, covers an area of 82,400 km², and had a total population of approximately 32.12 million in 2021. Chongqing is located between the middle reaches of the Yangtze River and the Sichuan Basin, with the landscape consisting of mountains, hills, and valleys. The terrain is characterized by large differences in elevation, with elevations ranging from 73.1 m to 2723.7 m. According to the Chongqing Municipal Bureau of Statistics (<https://tjj.cq.gov.cn/>, accessed on 8 July 2023), the urbanization rate in Chongqing reached 70.3% in 2021, and the city's annual GDP was CNY 2.7894 trillion. The mean temperature between 2015 and 2021 was 16–18 °C, and the mean precipitation during this period was approximately 1100–1400 mm (<https://ceidata.cei.cn/jsp/>, accessed on 12 July 2023).

Chongqing, China, ranked fourth in terms of the mean incidence rate of dysentery from 2015 to 2021, with a rate of 16.91 cases per 100,000 people. During this period, there were a total of 37,140 reported cases of dysentery in Chongqing. In addition, the overall incidence rate of dysentery in Chongqing was higher than the national average during this period (Figure 1c). Therefore, given Chongqing's unique geographical environment and the notably high incidence rate of dysentery, it is of paramount importance to investigate the dysentery incidence rate in Chongqing. Figure 1b also shows that the incidence rate of dysentery is relatively high in the areas surrounding the main city of Chongqing and expands northeastward. Moreover, Chengkou, located in the northernmost part of Chongqing, is also a high-risk area.

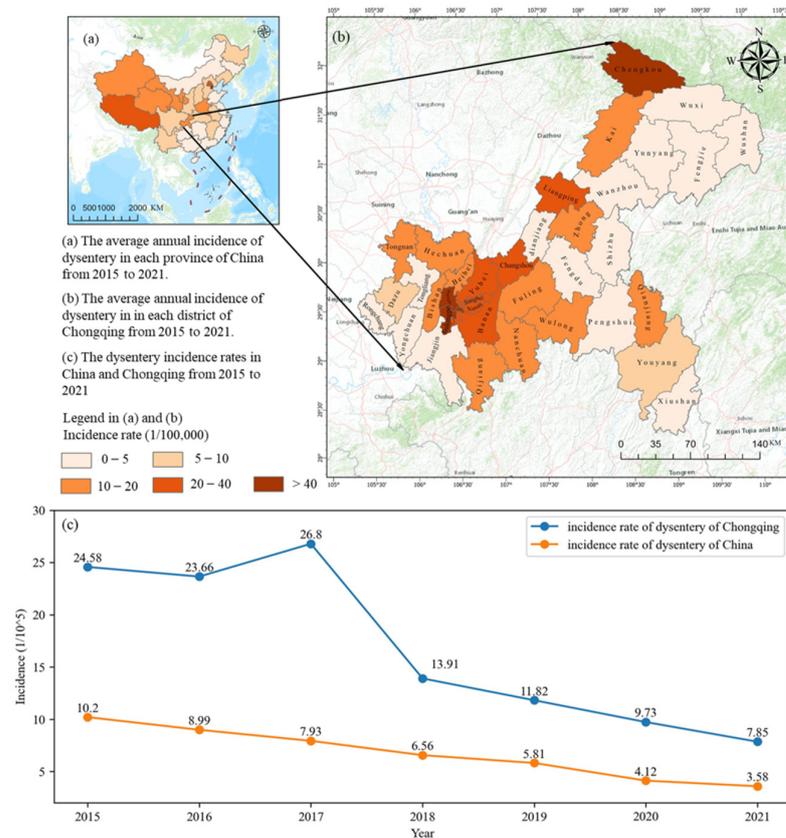


Figure 1. Incidence rate of dysentery in Chongqing.

2.2. Materials Sources and Processing

In this study, the incidence rates of dysentery and other auxiliary data such as socioeconomic factor data, meteorological factor data, topographic factor data, and air quality factor data in Chongqing from 2015 to 2021 were collected. The details are presented in Table 1.

Table 1. Datasets used in this study.

Name	Format	Spatial Resolution	Temporal Resolution	Data Source
Incidence rate data of dysentery	Text	/	Year	Chongqing Municipal Health Commission
Nighttime light dataset	Grid	1 km	Year	HARVARD Dataverse
Population	Grid	1 km	Year	LandScan
NDVI	Grid	0.01745° (~1 km)	Month	National Earth System Science Data Center
PM _{2.5}	Grid	0.01745° (~1 km)	Year	
PM ₁₀	Grid	0.01745° (~1 km)	Year	
Temperature	Grid	0.01745° (~1 km)	Month	
Precipitation	Grid	0.01745° (~1 km)	Month	National Tibetan Plateau Data Center
Relative humidity	Grid	1 km	Month	National Earth System Science Data Center/National Climatic Data Center
DEM	Grid	12.5 m	/	The Earth Science Data Systems

2.2.1. Incidence Rate of Dysentery in Chongqing

The incidence rate data of dysentery for all 38 districts in Chongqing from 2015 to 2021 were obtained from the website of the Chongqing Municipal Health Commission (<https://wsjkw.cq.gov.cn/>, accessed on 8 July 2023). In total, we collected 266 samples of incidence rate data.

2.2.2. Socioeconomic Factor Data

The Nighttime light (NTL) dataset (2015–2021) was obtained from improved time-series DMSP-OLS-like data product in China by integrating the DMSP-OLS and SNPP-VIIRS data [33] published in HARVARD Dataverse (<https://dataverse.harvard.edu/>, accessed on 8 July 2023). The spatial resolution of these data is 1 km.

The population raster dataset covering the years 2015 to 2021 was obtained from the open-source population mapping product called the LandScan Global population database, which is maintained by the Oak Ridge National Laboratory (<https://landscan.ornl.gov/>, accessed on 8 July 2023). The spatial resolution of the raw data is 0.01745 arc-degrees.

Monthly normalized difference vegetation index (NDVI) datasets were obtained from the National Earth System Science Data Center, National Science & Technology Infrastructure of China (<http://www.geodata.cn>, accessed on 8 July 2023). Their original resolution is approximately 0.01745 arc-degrees.

2.2.3. Meteorological Factor Data

Monthly mean temperature grid datasets [34–38] with a resolution of 0.01745 arc-degrees were collected from the National Tibetan Plateau Data Center (<https://data.tpdc.ac.cn/>, accessed on 4 July 2023) in China and spanned the period from 2015 to 2021.

Monthly precipitation datasets [38] covering China from 2015 to 2021 were obtained from the National Earth System Science Data Center, National Science & Technology Infrastructure of China. Their original resolution is approximately 0.01745 arc-degrees.

A monthly relative humidity gridded dataset covering the period from 2015 to 2020 with a spatial resolution of 1 km was collected from the National Earth System Science Data Center, National Science & Technology Infrastructure of China. Due to the lack of relative humidity raster data for 2021, air temperature and dew point temperature data of Chinese

regional meteorological stations through the FTP server (<ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-lite/>, accessed on 4 July 2023) published by the National Climatic Data Center (NCDC) were collected. The relative humidity was calculated from air temperature and dew point temperature by Equation (1) [39]. Afterward, kriging interpolation was performed on the relative humidity data recorded at regional stations in China to obtain raster data with a resolution of 1 km.

$$RH = 100 \times \frac{e^{(17.62 \times T_d)/(T_d + 243.12)}}{e^{(17.62 \times T)/(T + 243.12)}} \quad (1)$$

where RH represents relative humidity, T_d represents the dew point temperature, and T represents the air temperature.

2.2.4. Topographic Factor Data

The 12.5 m digital elevation model (DEM) data in Chongqing were obtained from the ALOS PALSAR Dataset of The Earth Science Data Systems (ESDS) (<https://search.asf.alaska.edu/>, accessed on 4 July 2023) in Chongqing. We proceeded to resample these data, changing the original resolution of 12.5 m to 1 km. Based on these resampled data, the surface slope was calculated.

2.2.5. Air Quality Factor Data

Yearly $PM_{2.5}$ and PM_{10} datasets covering China from 2015 to 2021 at a 1 km resolution were obtained from the National Earth System Science Data Center, National Science & Technology Infrastructure of China. The original resolution of these data is approximately 0.01745 arc-degrees.

2.2.6. Data Integration

To ensure the smooth operation of the model, the format of all covariates should be standardized. Therefore, NTL data were used as the basis to resample and correct all other grid data, including the DEM, slope, NDVI, $PM_{2.5}$, PM_{10} , precipitation, temperature, and relative humidity data, so that all data had the same spatial resolution (1 km), spatial extent, and row/column numbers. The NDVI, temperature, relative humidity, and precipitation data were all monthly data. However, this study was focused on an annual timescale, so these data needed to be aggregated on a yearly basis. Specifically, the maximum NDVI value across the 12 months was taken as the annual NDVI value, the monthly precipitation values were summed up to obtain the annual cumulative precipitation value, and the mean relative humidity and temperature across the 12 months were calculated to obtain the annual relative humidity and temperature values, respectively. This aggregation process enabled us to analyze the annual trends and patterns in these environmental variables.

2.3. Methods

To explore the nonlinear relationships between the dysentery incidence rate and the covariates, an IGBDT model was established to map the incidence rate of dysentery to a 1 km grid scale. First, the mean values of the covariates (mean NTL, mean NDVI, mean population, mean DEM, mean slope, mean cumulative precipitation, mean temperature, annual maximum temperature, annual minimum temperature, mean relative humidity, mean night light, and mean $PM_{2.5}$ and PM_{10}) in each district were used as the explanatory variables, and the annual incidence rate of dysentery in each district was used as the response variable to train the model. Due to the large gaps in the annual dysentery incidence rate among districts, a logarithmic transformation was performed to make the distribution of data more symmetrical and reduce the impacts of extreme values on the model. Then, the covariate data of Chongqing from 2015 to 2021 were input into the IGBDT model in pixels to predict the incidence rate of dysentery (anti-logarithm is required) at the pixel level at a 1 km resolution. The specific workflow is shown in Figure 2.

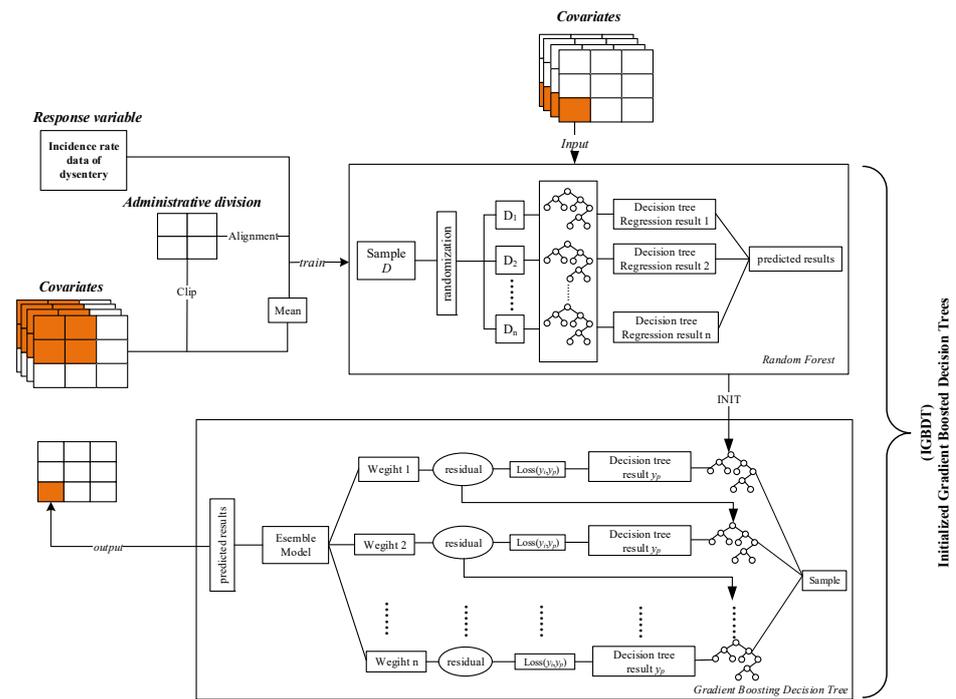


Figure 2. The workflow used to map the dysentery incidence rate to a fine-grained grid scale.

IGBDT is a method in which trained RF prediction results are used to initialize the GBDT, and the final boosted learner is added to the initial RF results. The specific algorithm process is described as follows. First, an RF is trained on the training data to obtain the initial prediction results. Then, the residuals of the initial prediction results are calculated and used as the target variables to train the GBDT. In each iteration of the GBDT, a new decision tree is trained to fit the residuals, and the results of the new decision tree are added to the initial prediction results of the RF. This process is repeated until the desired number of iterations is reached or until the residuals are no longer significantly reduced. The specific algorithm is shown in Steps (1)–(3).

Step 1. Initialize the weak learner:

$$f_0(x) = \operatorname{arccmin}_c \sum_{i=1}^n L(y_i, RF(x_i)) \quad (2)$$

where L is the loss function.

Step 2. Perform the iterations: where m is the number of weak learners.

$m \in \{1, 2, \dots, M\}$:

- (a) Calculate the negative gradient for each sample (x_i, y_i) ($i \in \{1, 2, \dots, n\}$). The obtained residual r_{im} is taken as the new true value of the sample, and (x_i, r_{im}) ($i \in \{1, 2, \dots, n\}$) is obtained as the training data of the next tree, then a new regression tree $f_m(x)$ is obtained along with its corresponding R_{jm} ($j \in \{1, 2, \dots, J\}$). J is the number of leaf nodes in the regression tree.

$$r_{i,m} = -\frac{\partial L(y_i, f_{m-1}(x_i))}{\partial f_{m-1}(x_i)} \quad (3)$$

- (b) Calculate the best fitting value for the leaf area J .

$$t_{jm} = \operatorname{arccmin}_t \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + t) \quad (4)$$

(c) Update the strong learner.

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^J t_{jm} I(x \in R_{jm}) \quad (5)$$

where I is a function. If the sample falls on the node, then $I = 1$; otherwise, $I = 0$.

Step 3. Obtain the final learner:

$$f(x) = f_0(x) + \sum_{m=1}^M \sum_{j=1}^J t_{jm} I(x \in R_{jm}) \quad (6)$$

3. Results

3.1. Evaluating the Quality of the Model

In this article, we divided the dataset into 80% for the training set and 20% for the testing set and utilized cross-validation in place of a validation set to evaluate the model. First, an RF model ($n = 860$, $\text{max_depth} = 12$, $\text{max_features} = 9$, $\text{min_samples_split} = 22$, $\text{min_samples_leaf} = 14$, and $\text{max_leaf_nodes} = 847$) was constructed using a Bayesian hyperparameter optimization training approach. The MAE, RMSE, and R^2 values of the testing set were $5.9345 (1/10^5)$, $8.1928 (1/10^5)$, and 0.7224 , respectively.

Subsequently, the previously trained RF model was used to initialize the GBDT model, resulting in the IGBDT model ($\text{init} = \text{RF}$, $\text{loss} = \text{'huber'}$, $\text{n_estimators} = 810$, $\text{learning_rate} = 0.015$, $\text{max_depth} = 8$, $\text{max_features} = 10$, $\text{min_samples_split} = 32$ and $\text{min_samples_leaf} = 4$), which was trained by Bayesian hyperparameter optimization. Figure 3 shows the specific performance of the IGBDT model. The MAE, RMSE, and R^2 values of the testing set were $4.7522 (1/10^5)$, $6.5265 (1/10^5)$, and 0.8239 . Compared to the RF, the R^2 value of the IGBDT increased by 0.1015 , and the MAE and RMSE values decreased by $1.1823 (1/10^5)$ and $1.6663 (1/10^5)$, respectively, in the testing set. The MAE, RMSE, and R^2 values obtained through five-fold cross-validation were $5.0039 (1/10^5)$, $7.346 (1/10^5)$, and 0.78 , respectively.

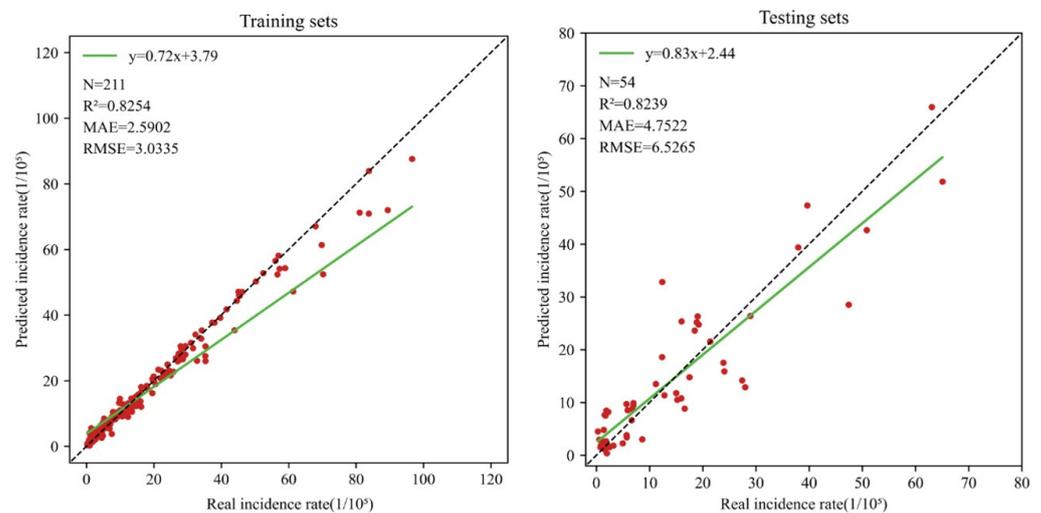


Figure 3. IGBDT model. The black dashed line is the perfect fitting line with R^2 equal to 1, red dots represent data distribution points, and the green solid line is the data fitting line.

3.2. Dysentery Incidence Rate Grained Scale Product (1 km)

The IGBDT model was used to map the incidence rate of dysentery in Chongqing to the pixel level (1 km). To reflect the spatial distribution of the incidence rate of dysentery in 1 km grained scale products, the mean incidence rate from 2015 to 2021 was calculated for grained scale products, as shown in Figure 4. The spatial distribution of the mean incidence rate was consistent with the distribution trend in 2019. Using this product, it is possible to understand the spatial distribution of dysentery incidence in the regions of Chongqing at a

relatively fine scale, not only at the scale of statistical units. The incidence rate of dysentery is relatively high in the main urban area of Chongqing and Chengkou, which is located in the northernmost part of Chongqing. High dysentery incidence rates can also be observed in the northwestern and southwestern of Chongqing. Additionally, there is a clear zone with high dysentery incidence rates along the Yangtze River basin.

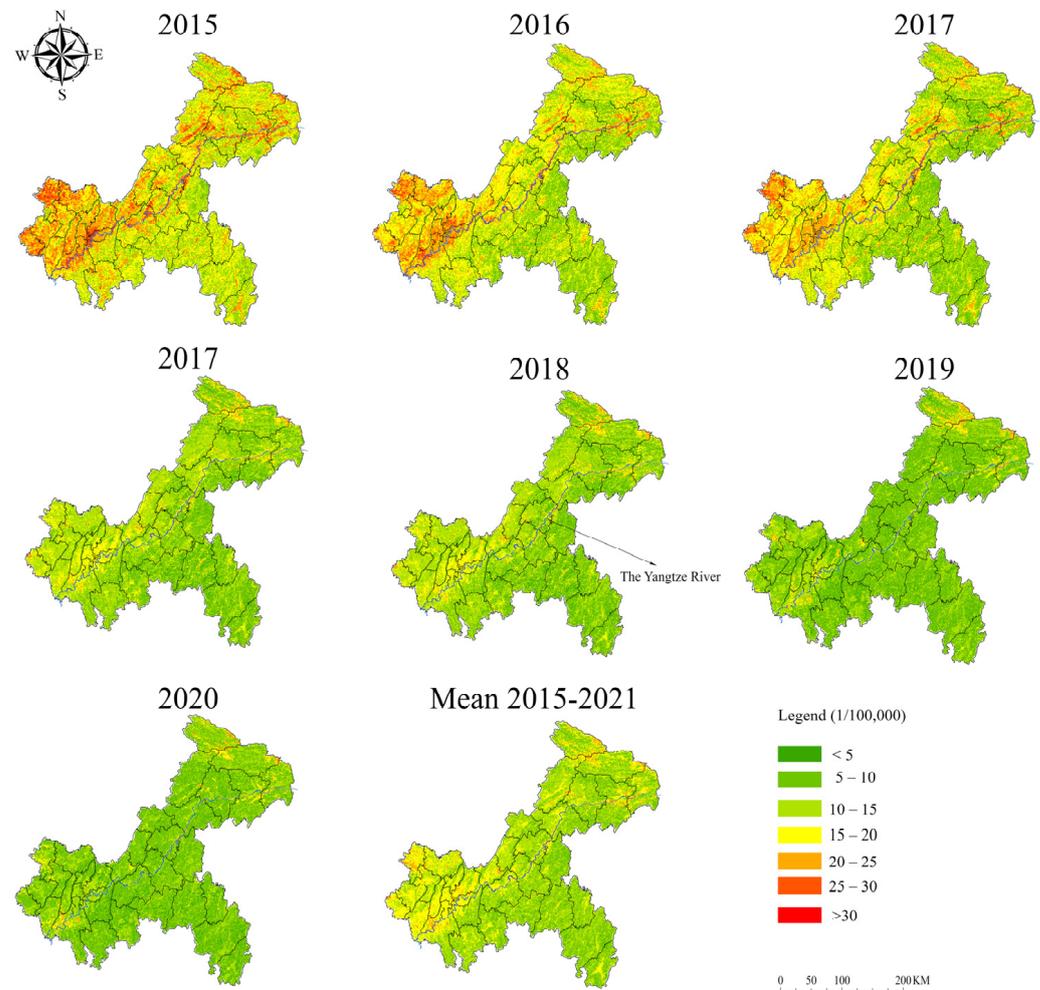
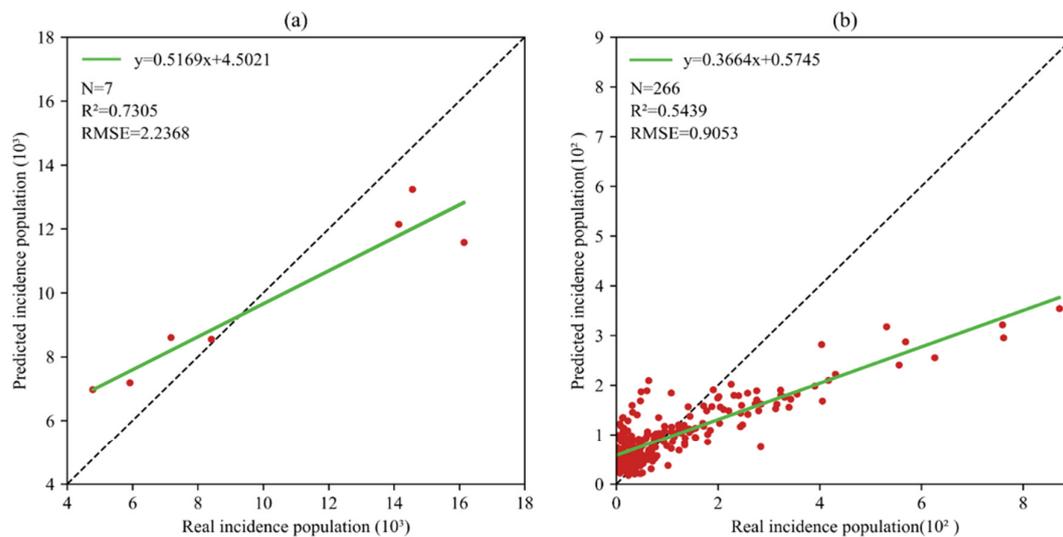


Figure 4. Grained scale products of the dysentery incidence rate (1 km) in Chongqing from 2015 to 2021.

A regression evaluation was carried out on the population data multiplied by the simulated incidence rates and actual incidence rate data in Chongqing, as presented in Figure 5. Figure 5a illustrates the regression analysis outcome of the actual and simulated incidence rate numbers for the entire city from 2015 to 2021, resulting in a correlation coefficient of 0.7360. Figure 5b displays the regression analysis results of the actual and simulated incidence rate numbers for each district in the city from 2015 to 2021, with a correlation coefficient of 0.5439. These results suggest that the model is capable of predicting the incidence rate to a certain extent and can capture some factors influencing dysentery incidence.



- (a) The regression analysis outcome of the actual and simulated incidence rate numbers for the entire city from 2015 to 2021.
- (b) The regression analysis results of the actual and simulated incidence rate numbers for each district in the city from 2015 to 2021.

Figure 5. Evaluation of morbidity grained scale products. The black dashed line is the perfect fitting line with R^2 equal to 1, red dots represent data distribution points, and the green solid line is the data fitting line.

3.3. Covariate Importance and Correlation Analysis

Through the IGBDT model, the contribution of each characteristic element to the incidence rate can be obtained (Figure 6). Among the various factors analyzed in this study, the population, NTL, and NDVI had the highest impact weights among the socioeconomic factors, accounting for approximately 43.32% of the total impact weight, with the population accounting for 17.07%, NTL accounting for 14.18%, and NDVI accounting for 12.07%. Among the meteorological factors, the mean temperature accounted for 12.20%, minimum temperature accounted for 9.55%, and maximum temperature accounted for 3.46%. The impact weights of cumulative precipitation (4.88%) and relative humidity (3.38%) were relatively small, with meteorological factors accounting for a total of 33.47%. The impacts of air quality factors, specifically PM_{10} and $PM_{2.5}$, were relatively small, accounting for only 4.83%. Topographic factors, including the slope and DEM, accounted for a total of 14.42%. The impacts of annual features, as this study was based on time-series data from 2015 to 2021, account for 3.95% of the total impact weight. Overall, socioeconomic factors had the dominant impact on the dysentery incidence rate, followed by meteorological factors, with temperature being the most important meteorological factor, while relative humidity and cumulative precipitation had relatively small impacts. Topographic factors were more important than air quality factors, but neither reached a significant level of importance.

The grained scale products (1 km) of dysentery incidence rate were used to study the correlations between various features and the incidence rate of dysentery. The r (Pearson correlation coefficient [40]) value was calculated by computing the pixel values of each covariate at the corresponding location from 2015 to 2021. The results are shown in Figure 7. The r value cannot represent the importance of each factor or the incidence rate of dysentery but represents only the correlation of the trend between 2015 and 2021. In Chongqing, among the socioeconomic factors, NTL and NDVI were all negatively correlated with the incidence rate of dysentery, and the population was weakly positively correlated with the dysentery incidence rate. Among the meteorological factors, the minimum temperature and mean temperature were positively correlated with the dysentery incidence rate, while the maximum temperature was negatively correlated, contrary to the actual expectations. The cumulative precipitation and relative humidity were positively correlated with the

dysentery incidence rate. The air quality factors, PM₁₀ and PM_{2.5}, were both negatively correlated with the incidence rate of dysentery.

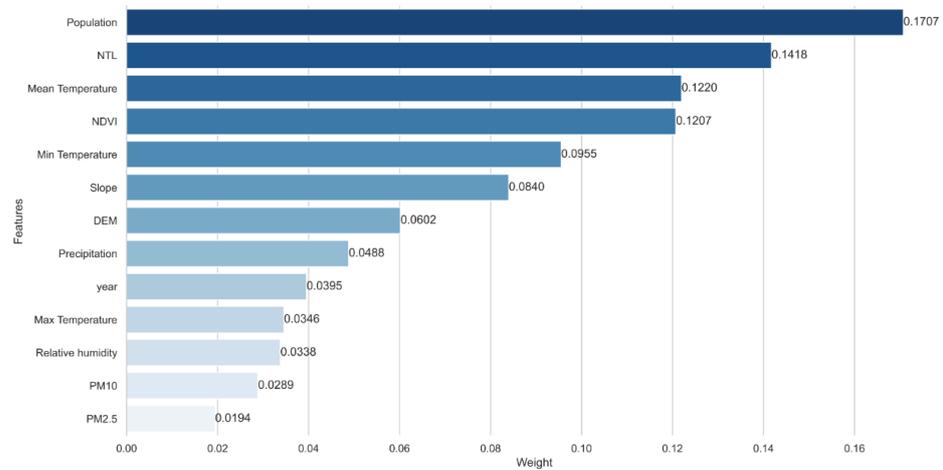


Figure 6. Feature importance.

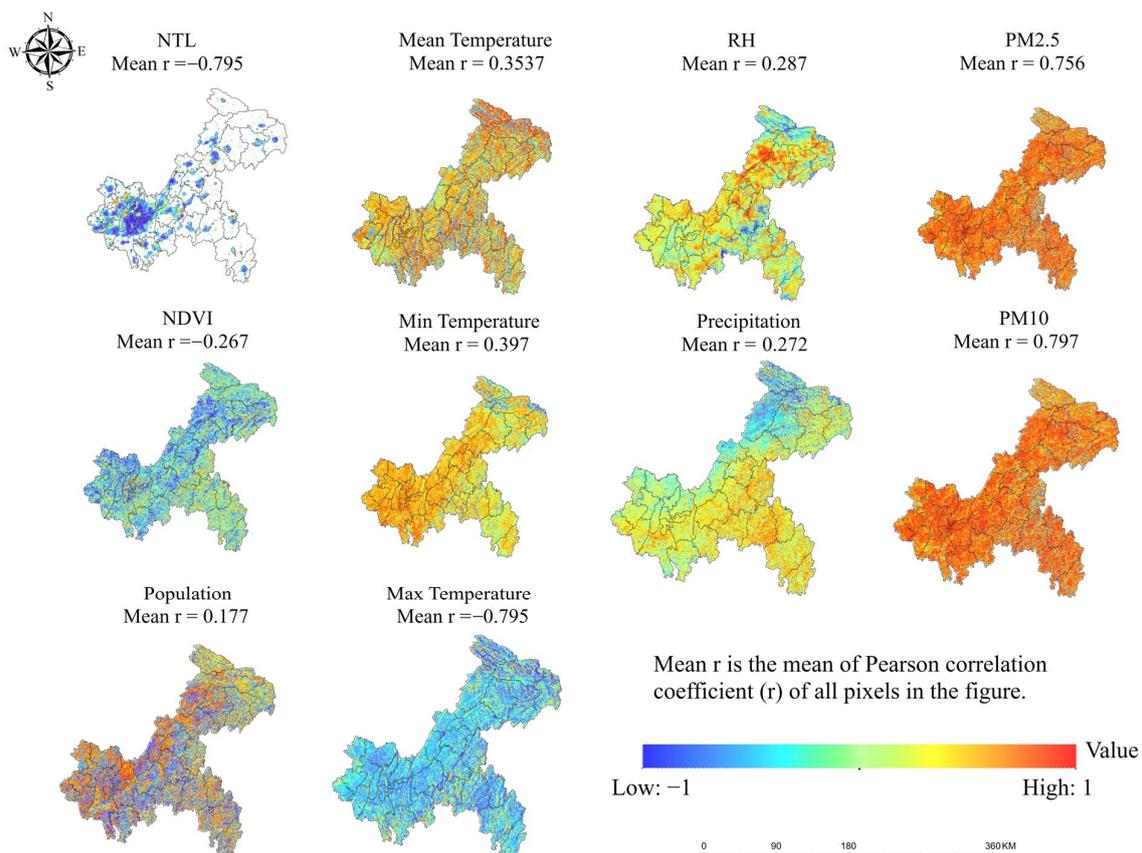


Figure 7. Correlation between each feature and the incidence rate of dysentery in Chongqing from 2015 to 2021.

The impacts of topographic factors on the dysentery incidence rate in Chongqing were investigated using the 2019 dysentery incidence rate grained scale product, which has the same spatial distribution trend as the mean incidence rate. The correlations between the incidence rate of dysentery and the DEM and slope were calculated using all the raster values of the dysentery incidence rate in 2019, and the resulting r values of the DEM and

slope were -0.2358 and -0.2371 , respectively. Our findings revealed that the impacts of topographic factors on the incidence rate of dysentery in Chongqing were negatively correlated, contrary to our initial expectations.

4. Discussion

4.1. Comparison with Other Models

We constructed the IGBDT, RF, GBDT, linear, and SVM models and obtained the optimal model through Bayesian hyperparameter optimization. The MAE, RMSE, and R^2 values of the IGBDT model in the testing set were best at 4.7024 ($1/10^5$), 6.2830 ($1/10^5$), and 0.8368, respectively. The MAE, RMSE, and R^2 values obtained by five-fold cross-validation were 5.0039 ($1/10^5$), 7.346 ($1/10^5$), and 0.78, respectively. Table 2 shows the performances of the different models. RF and GBDT perform better than linear and SVM. However, compared to GBDT, the performance of the RF was not ideal. The IGBDT model had a better performance than the RF and GBDT models. The R^2 values on the testing set improved by 0.1015 compared to the RF and by 0.0202 compared to the GBDT. Moreover, the MAE and RMSE values were better, demonstrating a higher accuracy, robustness, and generalization ability than the RF and GBDT models. Mohan et al. noted that the GBDT results were heavily influenced by initialization, while the RF was highly resistant to overfitting and served as an excellent optimization starting point. In addition, the IGBDT outperformed both RF and GBDT in terms of the RMSE values when M (the number of trees) was less than 1000. This conclusion was confirmed by our experiments, which indicated that providing GBDT with a better initialization can significantly improve its performance, allowing it to surpass the individual GBDT and RF models.

Table 2. Comparison of different models.

Model	MAE ($1/10^5$)	RMSE ($1/10^5$)	R^2
IGBDT	4.7024	6.2830	0.8368
RF	5.9345	8.1928	0.7224
GBDT	4.7260	6.6603	0.8166
Linear	9.1378	11.9004	0.4143
SVM	9.5465	11.487	0.4543

4.2. Comparison with Other Studies on the Influencing Factors of Dysentery

The transmission of dysentery may be influenced by various factors [17]. The incidence rate grained scale products (1 km) of dysentery were used to study the importance levels and correlations of socioeconomic factors, meteorological factors, air quality factors, and topographic factors on the dysentery incidence rate.

Significant negative correlations between the incidence rate of dysentery and the NTL and NDVI variables among socioeconomic factors were found, while a positive correlation was observed between the dysentery incidence rate and population density in this study. NTL satellite imagery is highly positively correlated with socioeconomic parameters, including urbanization, economic activity, and population [41–43]. In the first stage of urbanization (an urbanization rate of less than 77.59%), NDVI is positively correlated with the per-capita GDP [44]. The urbanization rate of Chongqing in 2021 was 70.3%, far from exceeding the first stage to the next stage. Therefore, the study of NTL and NDVI indirectly reflects that the relationship between the incidence rate of dysentery and socioeconomic development reflects a negative correlation; this finding is consistent with other research results [11,15,23,45,46]. Regarding the population density, the denser the population is, the more people the dysentery bacteria come into contact with within a certain time and space, and the more likely this situation is to have group reality. Therefore, we found a positive correlation between the population density and the dysentery incidence rate, consistent with previous research findings [6,8,23,24].

Meteorological factors such as temperature, relative humidity, and precipitation are considered to be among the main environmental predictors of the dysentery transmission

risk [47]. Previous studies have shown that temperature is a key meteorological factor affecting the incidence rate of dysentery [8,13,16,17,19,48]. For example, in a study performed around Beijing, the minimum, mean, and maximum temperatures were found to be positively correlated with the dysentery incidence rate [8]; the dysentery incidence rate in Jinan, China increased by 12%, and in Shenzhen, China, it increased by 16% with a 1 °C rise in the highest or lowest temperature [48]; in Peru, for every 1 °C increase in temperature, severe diarrhea in children increased by 8% [49]. The increases in temperature may lead to an increase in pathogen exposure, promote bacterial growth, and prolong the survival rate of bacteria in the environment and contaminated food [49]. In addition, when the temperature rises, some changes related to specific behavior in the population may occur, and such changes may increase the demand for drinking water and accelerate the spread of dysentery [50]. In this study, the minimum and mean temperatures were positively correlated with the incidence rate of dysentery, consistent with previous research conclusions. However, the maximum temperature was negatively correlated with the incidence rate, which was unexpected. A possible reason for this result is that high maximum temperatures may promote improved hygiene activities, such as increased hand washing and cleaning of food and water sources, thereby reducing the spread of dysentery. It is possible that other factors contributed to the decline in the incidence rate of dysentery during the study period, so specific, accurate research also needs to add more refined temperature factors to continue this discussion and research.

Positive correlations were observed between the dysentery incidence rate and cumulative precipitation and between the incidence rate and relative humidity. This was in line with the anticipated results based on previous studies. To date, some studies have explored the impacts of relative humidity and precipitation on the dysentery incidence rate, but the results are inconsistent. For example, some studies have shown positive correlations, such as studies performed in Northeast China, Hunan Province, and Beijing, where the incidence rates of dysentery were positively correlated with relative humidity and precipitation [5,13,14,47]. Similar findings have been reported in Taiwan, China [19], the Pacific islands [51], and Bangladesh [52]. Some studies have shown negative correlations, such as a study performed in Wuhan, which found that relative humidity and precipitation were negatively correlated with the risk of bacterial dysentery [17]. One study indicated that a lack of precipitation during the dry season would increase the incidence rate of dysentery in areas south of the Sahara Desert [18]. There are also studies that found no significant impact, such as a study conducted in two cities in North and South China, in which no significant correlation was found between relative humidity and precipitation and the incidence rate of dysentery [48,53].

Regarding air quality factors, we found that PM_{2.5} and PM₁₀ had positive correlations with the incidence rate of dysentery. Currently, no research has indicated any clear association between PM₁₀ or PM_{2.5} and dysentery. However, air pollution may affect the human immune system and health conditions, thereby increasing the risk of infectious diseases such as dysentery [25,54]. Therefore, the results of this study are in line with expectations, but studying the relationship between PM air pollution and the incidence rate of dysentery will require more detailed data.

The relationships between topographic factors and the incidence rate of bacillary dysentery are complex, with some studies indicating that topographic factors can lead to an increase in the dysentery incidence rate [15,46,55], while others have found that the impacts of topographic factors on the dysentery incidence rate are not significant [11]. These inconsistent results of existing studies may have been due to other factors, compared to topographic factors, being more important in determining the dysentery incidence rate, thus preventing generalizations. In this article, we found that the incidence rate of dysentery in Chongqing was negatively correlated with topographic factors overall. This was not consistent with the expected situation. The reason for this contradiction may be due to the complex terrain of Chongqing, which is low in the west and high in the east. However, the dysentery incidence rate is generally high in the west and low in the east. For

example, the terrain in the main urban area is flat, while the terrain in Chengkou is rugged and steep. The incidence rate of dysentery in both these places is high, while in the eastern part of the urban area, the terrain is also rugged and steep, but the incidence rate is low.

Our findings suggested high incidence rates of dysentery in the southwestern and northernmost parts of Chongqing. In the southwestern region, this high incidence rate is associated with factors such as high urbanization, a large population of migrants, a high population density, and a high per-capita GDP. This finding is consistent with previous studies that suggested that the population and economy are dominant factors affecting dysentery incidence rates [9,23]. On the other hand, Chengkou, located in the northernmost part of Chongqing with the least population (a permanent population of 198,000 in 2021) and per-capita GDP, has had a consistently high incidence rate since 2012 [9]. This may be due to the small population causing a relatively high disturbance in incidence rates. Although Wulong (with a permanent population of 356,500 in 2021) has a population one level higher than that of Chengkou, the incidence rate has not remained high. The specific reasons for these discrepancies need to be further investigated.

The obvious area of high dysentery incidence rates along the Yangtze River basin may be due to the low NDVI and DEM values in this area, leading to high dysentery incidence rates. Therefore, further optimization and research involving dysentery incidence rate prediction models should be conducted.

5. Conclusions

To fill the gap resulting from the lack of fine-grained scale dysentery incidence rate products at the grid scale, this study constructed an IGBDT hybrid machine learning model using socioeconomic factors (NTL, NDVI, and population), meteorological factors (precipitation, relative humidity and temperature), topographic factors (DEM and slope) and air quality factors (PM_{2.5} and PM₁₀) as explanatory variables, and the dysentery incidence rate as the output variable. The model was used to map annual statistical dysentery incidence rate data from various districts in Chongqing from 2015 to 2021 to the grid scale, thereby producing a dysentery incidence rate grid (1 km) product for Chongqing. The grained scale products of dysentery incidence rate were evaluated by using the total incidence of the population of Chongqing and each district in Chongqing with assessment R² results of 0.7369 and 0.5439, respectively. These results show that the model has a certain predictive effect. Comparing the IGBDT hybrid model with other models, such as the RF and GBDT models, it is proven that the IGBDT effect is better than that of either the RF or GBDT, individually. Then, the spatial distribution of the incidence of dysentery in Chongqing and the importance and correlation of its related factors were discussed. The results showed that the impacts of socioeconomic factors on the incidence of dysentery were the main factor, accounting for 43.32% of the impact, of which the NTL and NDVI showed negative correlations, and the population showed a positive correlation; meteorological factors accounted for 33.47%, of which the minimum temperature, mean temperature, precipitation, and relative humidity were positively correlated, while the maximum temperature was negatively correlated. However, the effects of topographic factors (the DEM and slope were negatively correlated) and air quality factors (PM_{2.5} and PM₁₀ were positively correlated) were relatively weak.

The dysentery incidence rate grained scale products for Chongqing (1 km) produced fill the gap left by the absence of detailed dysentery incidence rate products. It provides researchers and public health institutions with a more comprehensive foundation for their studies. This product allows researchers to delve deeper into the analysis of factors related to dysentery, revealing their significance and correlations. This, in turn, better supports the monitoring and control of dysentery. Furthermore, this grid product serves as a powerful tool for various sectors of society, assisting government and health institutions in accurately pinpointing potential high-risk areas of dysentery within densely populated and economically underdeveloped regions. It provides a more precise basis for resource allocation, enabling targeted governance and monitoring measures to reduce the spread

of dysentery. Compared to traditional methods based on administrative divisions, this fine-grained monitoring is more accurate and effective, holding promise for reducing the transmission of dysentery. In addition, the research approach and methods presented in this paper offer fresh ideas and references for the study of other diseases.

This study also has certain limitations. First, the dysentery incidence rate dataset included only 266 data from Chongqing between 2015 and 2021; this sample size is relatively small and may result in the overfitting of the model. Although the MAE, RMSE, and R^2 values of the testing set performed well in this study, when using the incidence population for verification, the R^2 value was only 0.56. The evaluation result R^2 value was not very high, further indicating that the model may have some limitations and cannot accurately predict the incidence of dysentery. Therefore, further research is needed to improve the accuracy of the model in the future to better cope with the prediction and control of dysentery. Second, due to the limited data considered in this study, the incidence of dysentery in Chongqing was mapped to the 1 km grid scale. In the future, obtaining data from a larger area would enable the production of a larger-scale grained scale product of dysentery incidence. Additionally, the feature selection in this study was limited. For example, hydrological factors greatly impact the spread of dysentery but were not considered here. Future work needs to include more influencing factors to predict and downscale the incidence of dysentery and to study its impact. Finally, in the application of the dysentery incidence grid, this study only analyzed the importance and correlations of influencing factors, and the results can represent only the study area. Further applications, such as refined prevention and control, should be conducted using dysentery incidence products obtained herein. Overall, this study has certain limitations, and future research should expand the dataset, consider more factors, and conduct more comprehensive application analyses.

Author Contributions: Jian Hao analyzed the data, constructed the IGBDT hybrid model, and wrote the manuscript. Jingwei Shen put forward the research objectives and ideas of the paper and was responsible for the planning and execution of research activities. All authors have read and agreed to the published version of the manuscript.

Funding: The research was supported by the Chongqing Social Science Planning Project (No. 2020PY28) and the Research Innovation Project of Southwest University (SWUS23080).

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: We would like to thank the editors and anonymous referees for their constructive comments.

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

References

1. Kotloff, K.L.; Riddle, M.S.; Platts-Mills, J.A.; Pavlinac, P.; Zaidi, A.K. Shigellosis. *Lancet* **2018**, *391*, 801–812. [[CrossRef](#)] [[PubMed](#)]
2. Holt, K.E.; Baker, S.; Weill, F.X.; Holmes, E.C.; Kitchen, A.; Yu, J.; Sangal, V.; Brown, D.J.; Coia, J.E.; Kim, D.W.; et al. Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nat. Genet.* **2012**, *44*, 1056–1059. [[CrossRef](#)]
3. Khezzani, B.; Baymakova, M.; Khechekhouche, E.A.; Ghezal, K.; Meziou, Z.; Brahim, A.B. Incidence rates of dysentery among humans in Lemghaier province, Algeria. *Germes* **2022**, *12*, 195–202. [[CrossRef](#)]
4. Xin, X.; Jia, J.; Hu, X.; Han, Y.; Liang, J.; Jiang, F. Association between floods and the risk of dysentery in China: A meta-analysis. *Int. J. Biometeorol.* **2021**, *65*, 1245–1253. [[CrossRef](#)]
5. Du, Z.; Zhang, J.; Lu, J.X.; Lu, L.P. Association between distribution of bacillary dysentery and meteorological factors in Beijing, 2004–2015. *Zhonghua Liu Xing Bing Xue Za Zhi* **2018**, *39*, 656–660. [[PubMed](#)]
6. Zhang, R.; Guo, Z.; Meng, Y.; Wang, S.; Li, S.; Niu, R.; Wang, Y.; Guo, Q.; Li, Y. Comparison of ARIMA and LSTM in Forecasting the Incidence of HFMD Combined and Uncombined with Exogenous Meteorological Variables in Ningbo, China. *Int. J. Environ. Res. Public Health* **2021**, *18*, 6174. [[CrossRef](#)] [[PubMed](#)]
7. Mbuyi, W.M.G.; Mandja, B.M.; Ilunga, B.K. Spatio-temporal dynamics of bacillary dysentery outbreaks in Democratic Republic of the Congo, 1999–2013. *Rev. D Epidemiol. Et De Sante Publique* **2021**, *69*, 1–6. [[CrossRef](#)] [[PubMed](#)]
8. Wang, L.; Xu, C.; Xiao, G.; Qiao, J.; Zhang, C. Spatial heterogeneity of bacillary dysentery and the impact of temperature in the Beijing-Tianjin-Hebei region of China. *Int. J. Biometeorol.* **2021**, *65*, 1919–1927. [[CrossRef](#)] [[PubMed](#)]

9. Zhang, P.; Zhang, J.; Chang, Z.R.; Li, Z.J. Temporal-spatial analysis of bacillary dysentery in the Three Gorges Area of China, 2005–2016. *Zhonghua Liu Xing Bing Xue Za Zhi* **2018**, *39*, 47–53.
10. Bahrami, G.; Sajadi, H.; Rafiee, H.; Norouzi, M.; Shakiba, A. Prediction of the Impacts of Climate Change on the Geographical Distribution of Dysentery in Iran. *Chin. J. Urban Environ. Stud.* **2022**, *10*, 2250018. [[CrossRef](#)]
11. Zuo, S.; Yang, L.; Dou, P.; Ho, H.C.; Dai, S.; Ma, W.; Ren, Y.; Huang, C. The direct and interactive impacts of hydrological factors on bacillary dysentery across different geographical regions in central China. *Sci. Total Environ.* **2021**, *764*, 144609. [[CrossRef](#)] [[PubMed](#)]
12. Ma, Y.; Wen, T.; Xing, D.; Zhang, Y. Associations between floods and bacillary dysentery cases in main urban areas of Chongqing, China, 2005–2016: A retrospective study. *Environ. Health Prev. Med.* **2021**, *26*, 1–9. [[CrossRef](#)]
13. Xu, C.; Xiao, G.; Wang, J.; Zhang, X.; Liang, J. Spatiotemporal Risk of Bacillary Dysentery and Sensitivity to Meteorological Factors in Hunan Province, China. *Int. J. Environ. Res. Public Health* **2018**, *15*, 47. [[CrossRef](#)] [[PubMed](#)]
14. Li, Z.J.; Zhang, X.J.; Hou, X.X.; Xu, S.; Zhang, J.S.; Song, H.B.; Lin, H.L. Nonlinear and threshold of the association between meteorological factors and bacillary dysentery in Beijing, China. *Epidemiol. Infect. Gastrointest. Tract* **2015**, *143*, 3510–3519. [[CrossRef](#)]
15. Ma, Y.; Zhang, T.; Liu, L.; Lv, Q.; Yin, F. Spatio-Temporal Pattern and Socio-Economic Factors of Bacillary Dysentery at County Level in Sichuan Province, China. *Sci. Rep.* **2015**, *5*, 15264. [[CrossRef](#)] [[PubMed](#)]
16. Gao, L.; Ding, G.; Jiang, B.; Li, X.; Zhou, M.; Zhang, Y.; Liu, Q. Meteorological Variables and Bacillary Dysentery Cases in Changsha City, China. *Am. J. Trop. Med. Hyg.* **2014**, *90*, 697–704. [[CrossRef](#)] [[PubMed](#)]
17. Li, Z.; Wang, L.; Sun, W.; Hou, X.; Yang, H.; Sun, L.; Xu, S.; Sun, Q.; Zhang, J.; Song, H.; et al. Identifying high-risk areas of bacillary dysentery and associated meteorological factors in Wuhan, China. *Sci. Rep.* **2013**, *3*, 3239. [[CrossRef](#)]
18. Bandyopadhyay, S.; Kanji, S.; Wang, L. The impact of rainfall and temperature variation on diarrheal prevalence in Sub-Saharan Africa. *Appl. Geogr.* **2012**, *33*, 63–72. [[CrossRef](#)]
19. Chou, W.-C.; Wu, J.-L.; Wang, Y.-C.; Huang, H.; Sung, F.-C.; Chuang, C.-Y. Modeling the impact of climate variability on diarrhea-associated diseases in Taiwan (1996–2007). *Sci. Total Environ.* **2010**, *409*, 43–51. [[CrossRef](#)]
20. Wan, H.; Yoon, J.; Srikrishnan, V.; Daniel, B.; Judi, D. Landscape metrics regularly outperform other traditionally-used ancillary datasets in dasymetric mapping of population. *Comput. Environ. Urban Syst.* **2023**, *99*, 101899. [[CrossRef](#)]
21. Cheng, Z.; Wang, J.; Ge, Y. Mapping monthly population distribution and variation at 1-km resolution across China. *Int. J. Geogr. Inf. Sci.* **2022**, *36*, 1166–1184. [[CrossRef](#)]
22. Chen, Y.; Wu, G.; Ge, Y.; Xu, Z. Mapping Gridded Gross Domestic Product Distribution of China Using Deep Learning with Multiple Geospatial Big Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 1791–1802. [[CrossRef](#)]
23. Xiao, G.; Xu, C.; Wang, J.; Yang, D.; Wang, L. Spatial-temporal pattern and risk factor analysis of bacillary dysentery in the Beijing-Tianjin-Tangshan urban region of China. *BMC Public Health* **2014**, *14*, 998. [[CrossRef](#)] [[PubMed](#)]
24. Zhang, X.; Gu, X.; Wang, L.; Zhou, Y.; Huang, Z.; Xu, C.; Cheng, C. Spatiotemporal variations in the incidence of bacillary dysentery and long-term effects associated with meteorological and socioeconomic factors in China from 2013 to 2017. *Sci. Total Environ.* **2021**, *755*, 142626. [[CrossRef](#)] [[PubMed](#)]
25. Deryugina, T.; Heutel, G.; Miller, N.H.; Molitor, D.; Reif, J. The mortality and medical costs of air pollution: Evidence from changes in wind direction. *Am. Econ. Rev.* **2019**, *109*, 4178–4219. [[CrossRef](#)]
26. Han, X.; Zhang, W.; Cui, X.; Ma, H.; Liu, Y.; Zhang, X.; Zhao, X.; Li, S.; Ren, X. Effects of ambient temperature and air pollutants on bacillary dysentery from 2014 to 2017 in Lanzhou, China. Available online: <https://www.researchsquare.com/article/rs-4357/v2> (accessed on 8 July 2023).
27. Tu, W.; Liu, Z.; Du, Y.; Yi, J.; Liang, F.; Wang, N.; Qian, J.; Huang, S.; Wang, H. An ensemble method to generate high-resolution gridded population data for China from digital footprint and ancillary geospatial data. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102709. [[CrossRef](#)]
28. Huang, C.-J.; Kuo, P.-H. A Deep CNN-LSTM Model for Particulate Matter (PM_{2.5}) Forecasting in Smart Cities. *Sensors* **2018**, *18*, 2220. [[CrossRef](#)]
29. Mohan, A.; Chen, Z.; Weinberger, K. Web-Search Ranking with Initialized Gradient Boosted Regression Trees. *Proc. Learn. Rank. Chall. PMLR* **2011**, *14*, 77–89.
30. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
31. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
32. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [[CrossRef](#)]
33. Wu, Y.; Shi, K.; Chen, Z.; Liu, S.; Chang, Z. An Improved Time-Series DMSP-OLS-Like Data (1992–2021) in China by Integrating DMSP-OLS and SNPP-VIIRS. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [[CrossRef](#)]
34. Ding, Y.; Peng, S. Spatiotemporal Trends and Attribution of Drought across China from 1901–2100. *Sustainability* **2020**, *12*, 477. [[CrossRef](#)]
35. Peng, S.; Ding, Y.; Wen, Z.; Chen, Y.; Cao, Y.; Ren, J. Spatiotemporal change and trend analysis of potential evapotranspiration over the Loess Plateau of China during 2011–2100. *Agric. For. Meteorol.* **2017**, *233*, 183–194. [[CrossRef](#)]
36. Peng, S.; Gang, C.; Cao, Y.; Chen, Y. Assessment of climate change trends over the Loess Plateau in China from 1901 to 2100. *Int. J. Climatol.* **2018**, *38*, 2250–2264. [[CrossRef](#)]

37. Peng, S.Z. *1-km Monthly Mean Temperature Dataset for China (1901–2021)*; National Tibetan Plateau Data Center: Beijing, China, 2019.
38. Peng, S.; Ding, Y.; Liu, W.; Li, Z. 1 km monthly temperature and precipitation dataset for China from 1901 to 2017. *Earth Syst. Sci. Data* **2019**, *11*, 1931–1946. [[CrossRef](#)]
39. Wexler, A. Vapor pressure formulation for water in range 0 to 100 C. A revision. *J. Res. Natl. Bur. Stand. Sect. A Phys. Chem.* **1976**, *80*, 775. [[CrossRef](#)]
40. Cohen, I.; Huang, Y.; Chen, J.; Benesty, J. *Pearson Correlation Coefficient*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4.
41. Keola, S.; Andersson, M.; Hall, O. Monitoring Economic Development from Space: Using Nighttime Light and Land Cover Data to Measure Economic Growth. *World Dev.* **2015**, *66*, 322–334. [[CrossRef](#)]
42. Shi, K.; Chen, Y.; Yu, B.; Xu, T.; Yang, C.; Li, L.; Huang, C.; Chen, Z.; Liu, R.; Wu, J. Detecting spatiotemporal dynamics of global electric power consumption using DMSP-OLS nighttime stable light data. *Appl. Energy* **2016**, *184*, 450–463. [[CrossRef](#)]
43. Sutton, P.C.; Elvidge, C.D.; Ghosh, T. Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. *Int. J. Ecol. Econ. Stat.* **2007**, *8*, 5–21.
44. Liu, J.; Zhang, Z.; Xu, X.; Kuang, W.; Zhou, W.; Zhang, S.; Li, R.; Yan, C.; Yu, D.; Wu, S.; et al. Spatial patterns and driving forces of land use change in China during the early 21st century. *J. Geogr. Sci.* **2010**, *20*, 483–494. [[CrossRef](#)]
45. Ferrer, S.R.; Strina, A.; Jesus, S.R.; Ribeiro, H.C.; Cairncross, S.; Rodrigues, L.C.; Barreto, M.L. A hierarchical model for studying risk factors for childhood diarrhoea: A case–control study in a middle-income country. *Int. J. Epidemiol.* **2008**, *37*, 805–815. [[CrossRef](#)] [[PubMed](#)]
46. Zhang, H.; Si, Y.; Wang, X.; Gong, P. Environmental Drivers and Predicted Risk of Bacillary Dysentery in Southwest China. *Int. J. Environ. Res. Public Health* **2017**, *14*, 782. [[CrossRef](#)]
47. Huang, D.; Guan, P.; Guo, J.; Wang, P.; Zhou, B. Investigating the effects of climate variations on bacillary dysentery incidence in northeast China using ridge regression and hierarchical cluster analysis. *BMC Infect. Dis.* **2008**, *8*, 130. [[CrossRef](#)]
48. Zhang, Y.; Bi, P.; Hiller, J.E.; Sun, Y.; Ryan, P. Climate variations and bacillary dysentery in northern and southern cities of China. *J. Infect.* **2007**, *55*, 194–200. [[CrossRef](#)] [[PubMed](#)]
49. Checkley, W.; Epstein, L.D.; Gilman, R.H.; Figueroa, D.; Cama, R.I.; Patz, J.A.; Black, R.E. Effects of El Niño and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. *Lancet* **2000**, *355*, 442–450. [[CrossRef](#)] [[PubMed](#)]
50. Black, R.E.; Lanata, C.F. Epidemiology of diarrheal diseases in developing countries. *Infect. Gastrointest. Tract* **1995**, *2*, 11–29.
51. Singh, R.B.; Hales, S.; De Wet, N.; Raj, R.; Hearnden, M.; Weinstein, P. The influence of climate variation and change on diarrheal disease in the Pacific Islands. *Environ. Health Perspect.* **2001**, *109*, 155–159. [[CrossRef](#)]
52. Hashizume, M.; Ben Armstrong, B.; Hajat, S.; Wagatsuma, Y.; Faruque, A.S.; Hayashi, T.; Sack, D.A. Association between climate variability and hospital visits for non-cholera diarrhoea in Bangladesh: Effects and vulnerable groups. *Int. J. Epidemiol.* **2007**, *36*, 1030–1037. [[CrossRef](#)]
53. Zhang, Y.; Bi, P.; Hiller, J.E. Weather and the transmission of bacillary dysentery in Jinan, northern China: A time-series analysis. *Public Health Rep.* **2008**, *123*, 61–66. [[CrossRef](#)]
54. Di, Q.; Wang, Y.; Zanobetti, A.; Wang, Y.; Koutrakis, P.; Choirat, C.; Dominici, F.; Schwartz, J.D. Air pollution and mortality in the Medicare population. *N. Engl. J. Med.* **2017**, *376*, 2513–2522. [[CrossRef](#)] [[PubMed](#)]
55. Zhang, H.; Si, Y.; Wang, X.; Gong, P. Patterns of bacillary dysentery in China, 2005–2010. *Int. J. Environ. Res. Public Health* **2016**, *13*, 164. [[CrossRef](#)] [[PubMed](#)]

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.