

Article

A High-Resolution Dataset for Arabica Coffee Distribution in Yunnan, Southwestern China

Hongyu Shan ^{1,2,3}, Tao Ye ^{1,2,3,*}, Zhe Chen ^{4,5} , Wenzhi Zhao ³ , Xuehong Chen ³ and Hao Sun ^{4,5}

¹ State Key Laboratory of Earth Surface Processes and Disaster Risk Reduction (ESPD RR), Beijing Normal University, Beijing 100875, China

² Ministry of Emergency Management, Ministry of Education, Academy of Disaster Reduction and Emergency Management, Beijing 100875, China

³ Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

⁴ State Key Laboratory for Vegetation Structure, Functions and Construction, Yunnan University, Kunming 650500, China

⁵ Yunnan Key Laboratory of Soil Erosion Prevention and Green Development, Institute of International Rivers and Eco-Security, Yunnan University, Kunming 650500, China

* Correspondence: yetao@bnu.edu.cn; Tel.: +86-10-58800268

Highlights

What are the main findings?

- A 10-m resolution Arabica coffee distribution dataset covering the main production areas of Yunnan, China, was developed based on Sentinel-2 imagery and terrain data, providing spatially explicit and up-to-date information in a complex mountainous environment.
- The study establishes an operational object-based mapping workflow integrating multi-seasonal spectral and topographic information, and identifies key seasonal spectral features that consistently contribute to reliable coffee discrimination.

What are the implications of the main findings?

- The proposed workflow offers a transferable and scalable framework for perennial crop mapping in heterogeneous mountainous regions, supporting regional agricultural monitoring and land management applications.
- The high-resolution coffee distribution dataset provides an essential spatial baseline for assessing land-use dynamics, ecological impacts, and the sustainability of coffee expansion in southwestern China.

Abstract

Coffee, as a perennial commodity crop, plays a crucial role in global agricultural markets, regional livelihoods, and poverty alleviation. Yunnan Province of China (21°8′–29°15′N) represents the northernmost coffee-growing region worldwide, and its production has gained increasing attention in international markets. However, the absence of a spatially explicit and high-resolution coffee distribution dataset has constrained environmental assessment, land-use analysis, and policy-making in this subtropical and marginal growing region. In this study, we developed the first 10 m resolution Arabica coffee distribution dataset for Yunnan Province for the year 2023 using Sentinel-2 optical imagery and Shuttle Radar Topographic Mission (SRTM) terrain data within the Google Earth Engine (GEE) platform. An object-based workflow was implemented to generate spatially coherent mapping units, followed by supervised classification to identify coffee plantations. The resulting map achieved an overall accuracy (OA) of 0.87, with user accuracy (UA), producer accuracy (PA), and F1 score of 0.90, 0.96, and 0.93 for the coffee class, demonstrating its reliability for regional-scale applications. Feature contribution analysis indicates that shortwave infrared



Academic Editors: Li Fang, Jian Yang and Yu Feng

Received: 6 January 2026

Revised: 8 March 2026

Accepted: 9 March 2026

Published: 19 March 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

(SWIR) and red-edge information, particularly during the dry season, plays an important role in coffee discrimination. These results enhance confidence in the ecological relevance and stability of the mapping framework. The proposed workflow provides a practical and transferable approach for perennial crop mapping in complex mountainous environments. More importantly, the generated high-resolution coffee distribution dataset establishes a spatial baseline for monitoring land-use dynamics, assessing ecological impacts, and supporting sustainable coffee development in southwestern China.

Keywords: coffee mapping; Yunnan; deep neural network; simple non-iterative clustering segmentation

1. Introduction

Accurate spatial distribution information of crops is essential for crop yield estimation, growth monitoring, disaster assessment, agricultural resource allocation, and the adjustment of planting structures [1–3]. Economic crops not only meet human food demands but also play a crucial role in promoting local economies and maintaining ecological balance [4]. Among these, coffee is one of the most valuable cash crops globally, ranking first in yield, output value, and consumption among the three major beverage crops [5]. It holds significant importance in the global agricultural market, the supply chain, and the livelihoods of millions of smallholder farmers [6]. Detailed, high-resolution maps of coffee distribution are essential for understanding the driving factors behind local biodiversity changes (e.g., deforestation), promoting sustainable development, improving supply chain transparency, and analyzing the interactions between coffee cultivation and climate change [7,8]. However, major coffee-producing regions worldwide generally lack publicly available, high-resolution (better than 30 m) coffee distribution maps, hindering further research and informed policy-making.

Currently, only a few coffee-growing regions, such as Indonesia, Mexico, and Vietnam, have access to high-resolution coffee distribution data [9–11]. For instance, Tridawati used pan-sharpened GeoEye-1, multi-temporal Sentinel-2, and digital elevation models (DEM) to map coffee distribution in parts of Mount Puntang, Indonesia [11]. Maskell generated the first 10 m resolution coffee map for Daklak Province, the primary coffee-growing region in Vietnam, distinguishing between shaded coffee, open-field coffee, and newly planted coffee, using Sentinel-1 and Sentinel-2 data along with terrain information [10]. Similarly, Escobar-López developed a dataset distinguishing between naturally shaded coffee and artificially planted shaded coffee by integrating monthly Sentinel-1/2 data, climate data, and terrain information during the dry season [9]. Despite these advancements, high-resolution coffee distribution data remains scarce even among the top five coffee-producing countries.

Most current remote sensing-based coffee mapping studies rely on supervised classification methods, particularly random forest algorithms, to identify coffee plantations using optical, radar, texture, seasonal, and terrain features derived from Sentinel-1 and Sentinel-2 data [9,10]. However, these methods face significant challenges due to the limited availability of ground control points [7]. Coffee is typically grown in steep, economically underdeveloped mountainous regions, making field data collection particularly difficult [10]. Furthermore, the fragmented nature of land parcels and the continuous cloud cover during the coffee-growing season exacerbate the challenges of accurately mapping coffee plantations [9]. The steep topography also weakens radar signals, reducing classification accuracy. Combining multi-source remote sensing data is considered an effective solution to these issues by integrating complementary information from optical

imagery, phenological patterns, and terrain features [7,12,13]. Given the altitude-dependent nature of coffee cultivation, terrain characteristics such as elevation and slope are particularly informative for classification [14,15]. The seasonal characteristics exhibited by coffee due to phenology or agricultural management measures also increase its differentiation from deciduous forests and evergreen broad-leaved forests, particularly between wet and dry seasons.

In addition to the selection of classification features, classification methods also have an impact on the accuracy of coffee remote sensing mapping. The targets of classification in coffee mapping are mainly pixel-based, which often leads to salt and pepper noise in the classification results due to the small size of coffee plots relative to other land cover types under terrain influence [16,17]. In contrast, object-based classification approaches segment pixels into meaningful objects based on spectral and spatial similarity, improving classification accuracy by leveraging spatial continuity and homogeneity [7,18]. Moreover, recent advances in deep learning algorithms and cloud computing platforms offer new opportunities to enhance the accuracy and efficiency of coffee mapping. Deep learning can automatically extract meaningful features from large datasets, while cloud platforms like Google Earth Engine (GEE) simplify the acquisition and processing of remote sensing images [19,20]. Despite their potential, these advanced methods have not been fully exploited for coffee mapping.

Yunnan Province accounts for 98% of China's coffee production and approximately 2% of the world's total output [21]. Most Yunnan coffee is grown between 21° and 26°N at elevations of 800–1500 m above sea level [22,23]. These geographic conditions give Yunnan coffee distinctive qualities [24] but also expose it to environmental risks such as frost and drought [25–27]. The unique ecological and agricultural characteristics of Yunnan's coffee industry offer valuable case studies for global coffee research and provide important opportunities for crop improvement and sustainable development [28]. As one of the region's main cash crops, coffee significantly contributes to local farmers' incomes. However, the absence of reliable, high-resolution coffee distribution data limits the ability of government agencies and researchers to conduct yield estimation, environmental monitoring, and policy development.

In this study, we developed an operational object-based mapping workflow to produce the first 10 m resolution coffee distribution dataset for Yunnan Province by integrating multi-source remote sensing data within the Google Earth Engine (GEE) platform. Spatially coherent image objects were generated using the Simple Non-Iterative Clustering (SNIC) algorithm to reduce within-class spectral variability and improve the representativeness of mapping units. Supervised classification was then conducted based on object-level spectral and topographic features, and model performance was evaluated to ensure the robustness of the resulting dataset. In addition, SHAP analysis was applied to assess feature contributions and enhance the transparency and interpretability of the mapping results. The resulting high-resolution coffee distribution map provides a spatially explicit baseline for monitoring coffee expansion, assessing land-use dynamics, and evaluating ecological and climatic impacts in this marginal growing region. By establishing a transferable and scalable mapping workflow, this study supports future agricultural monitoring and sustainable land management efforts in southwestern China.

2. Materials and Methods

2.1. Framework

In this study, we combined Sentinel-2 optimal data and SRTM topographic data to generate a high-resolution coffee distribution map of Yunnan. The methodology involves

four main steps (Figure 1): (1) data preparation, (2) object-based SNIC segmentation, (3) Deep Neural Network (DNN) classification based on GEE, and (4) accuracy assessment.

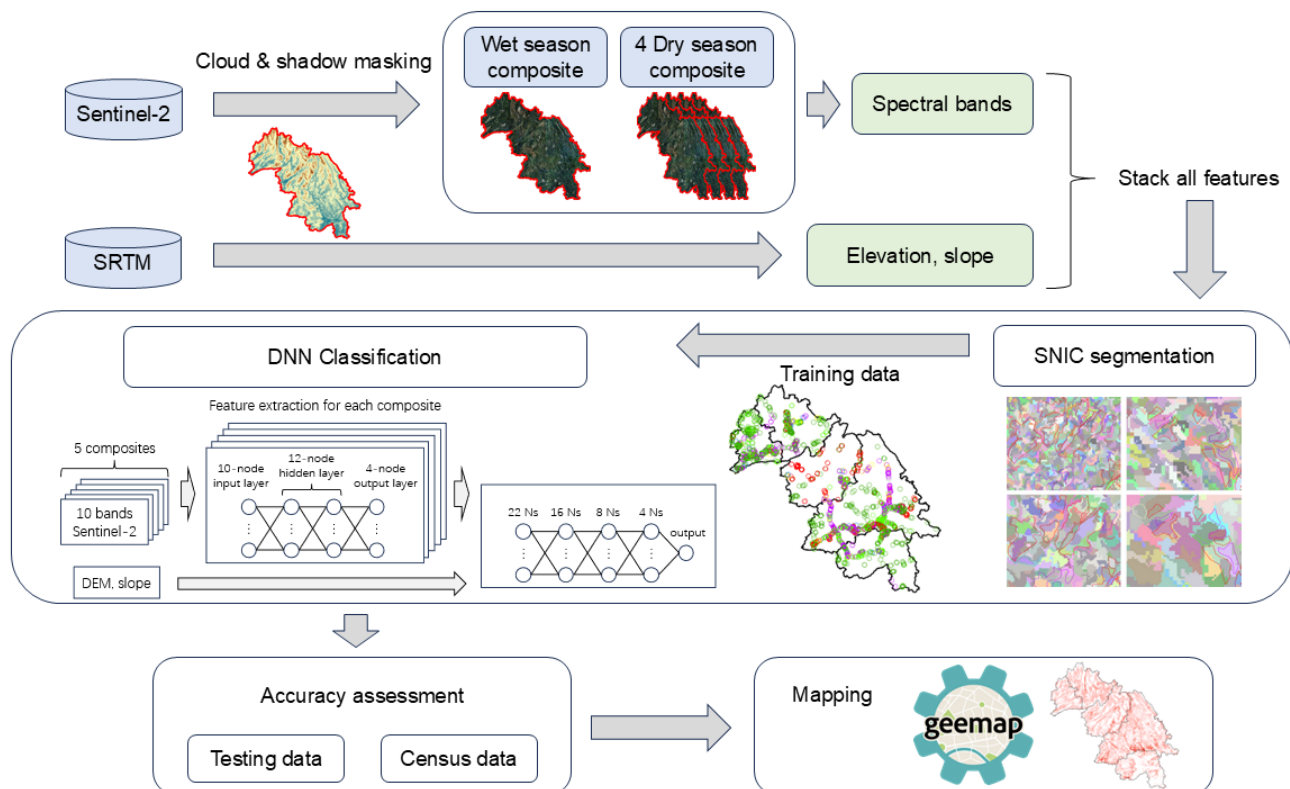


Figure 1. Framework for coffee mapping and validation.

2.2. Study Area

Yunnan is located in southwestern China ($97^{\circ}31'E$ – $106^{\circ}11'E$, $21^{\circ}8'N$ – $29^{\circ}15'N$) and is characterized by highly diverse terrain types, primarily consisting of plateau mountains (Figure 2). The elevation ranges from 225 m to 6740 m above sea level. Coffee cultivation is concentrated in five key administrative regions: Dehong Prefecture, Baoshan City, Lincang City, Pu'er City, and Xishuangbanna Prefecture, situated in the southwestern part of the province. These regions exhibit distinct vertical climatic characteristics, with temperature decreasing with increasing altitude. The coffee-growing areas experience a subtropical plateau monsoon climate [23], with an average temperature of $19\sim 22^{\circ}C$ in the hottest month (July) and $6\sim 8^{\circ}C$ in the coldest month (January). The wet season occurs from April to October, accounting for over 85% of the annual precipitation, while the dry season spans from November to March, contributing only 15% of the annual rainfall [29]. The county-level administrative divisions of the research area are shown in Figure S1.

The primary coffee variety cultivated in Yunnan is Arabica (*Coffea arabica*), and the predominant planting method is open-canopy cultivation on large, high-altitude slopes. Some coffee plots incorporate scattered shade trees, such as macadamia (*Macadamia ternifolia* F. Muell.), while only a small fraction of the surveyed plots employ intercropping or fully shaded planting systems. The phenological cycle of coffee is closely linked to the region's distinct dry and wet seasons [30]. Flowering typically begins in March as temperatures rise. Following the onset of the rainy season in May, the flowers wither and fruit (coffee cherries) begin to develop. During the dry season starting in November, the cherries ripen and change color from green to yellow, orange, or red, marking the harvesting period. After harvest, the coffee plants enter a dormancy and nutrient accumulation phase.

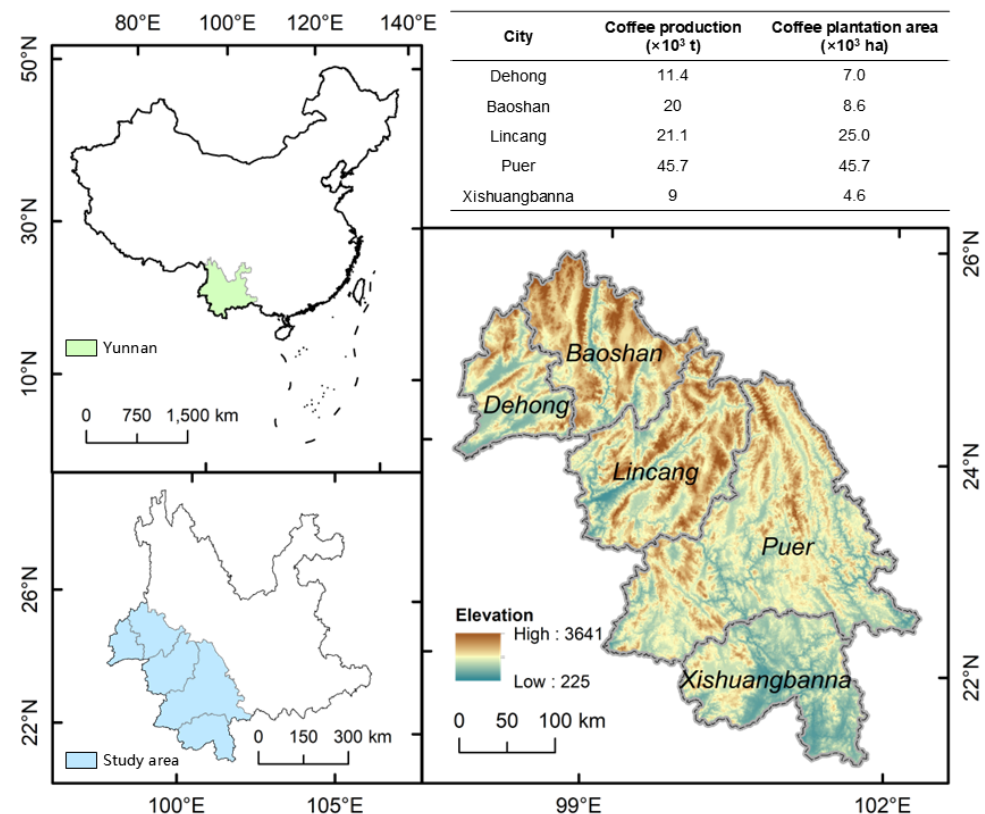


Figure 2. Location of main coffee growing area in Yunnan province, China.

2.3. Data

2.3.1. Remote Sensing Data

We obtained Sentinel-2 (S2) optical data from April 2023 to March 2024 and SRTM 30 m DEM data from the GEE [20]. Sentinel-2 is a wide-swath, high-resolution multispectral imaging mission with a global revisit frequency of five days [31]. Its Multispectral Instrument (MSI) captures 13 spectral bands: visible and near-infrared at 10 m, red edge and short-wave infrared (SWIR) at 20 m, and atmospheric correction bands at 60 m, providing comprehensive data for monitoring vegetation, soil, and water changes [32].

For this study, we used the S2 Level-2A Surface Reflectance product (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED, accessed on 1 May 2024), which has been atmospherically corrected using the Sen2Cor algorithm [33,34]. To identify cloudy pixels, we employed the Sentinel-2 Cloud Probability (S2C) product (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_CLOUD_PROBABILITY, accessed on 1 May 2024), which provides a cloud probability value ranging from 0 to 100 at a 10 m resolution [35]. This product offers higher spatial accuracy and greater flexibility than the original QA60 band of the Sentinel-2 dataset. To mitigate the effects of cloud contamination, we applied a dual cloud masking approach using the QA60 band [36,37] and the S2C product, where pixels with cloud probabilities greater than 50% were classified as cloud-affected and excluded [34]. The bands (Table 1) we used included six 20 m resolution bands (red-edge bands, B5, B6, B7 and B8A; short-wave infrared bands, B11 and B12) and four 10 m resolution bands (blue, B2; green, B3; red, B4; and near-infrared, B8).

Table 1. Sentinel-2 band information.

Feature Name	Description	Feature Name	Description
Blue	Blue band	Re3	Red edge 3
Green	Green band	Nir	Near-infrared
Red	Red band	Re4	Red edge 4
Re1	Red edge 1	Swir1	Shortwave infrared 1
Re2	Red edge 2	Swir2	Shortwave infrared 2

To capture the phenological characteristics of coffee, we tested multiple temporal partitioning schemes to ensure adequate valid observations across the study area (Figure S2). During the dry season, image availability was sufficient to allow one complete image every month, with the exception of January and December where data gaps required merging imagery from both months. For the rainy season, however, cloud cover is too frequent to obtain any single cloud-free image for rainy-season months. Consequently, all rainy-season images were aggregated into a single composite image to achieve complete spatial coverage. Finally, we adopted a five-period segmentation approach that balances temporal coverage and phenological representation. Each coffee growth cycle was divided into the following periods: March (flowering onset), April to October (wet season, fruit development), November (fruit maturation), December to January (ripening and early dormancy), and February (late dormancy). We calculated the median value of the images within each period to reduce the influence of outliers and extreme values.

The SRTM project (https://developers.google.com/earth-engine/datasets/catalog/USGS_SRTMGL1_003, accessed on 1 May 2024) provides digital elevation data at a near-global scale, covering latitudes from 56°S to 60°N. From this dataset, we derived altitude and slope information to incorporate terrain effects into the classification process. Although aspect may influence local light and moisture conditions, field investigations revealed it does not have a decisive impact on coffee distribution; therefore, we excluded it from the predictive features.

2.3.2. Ground Samples

We conducted two field surveys in Yunnan Province during May and August 2023, each lasting approximately two weeks. Using the Ovital Map v10.4.2 (<https://www.ovital.com/>), we collected a total of 3695 ground-truth plot polygons with detailed category information (Table 2). These samples span a diverse range of terrains, vegetation types, and land-use patterns, ensuring comprehensive representation of the study area's ecological and agricultural variability. To improve the accuracy of our classification, we collected more samples in areas where coffee is grown, and paid special attention to collecting samples of both coffee and non-coffee land features that are easily confused with coffee. To avoid spatial autocorrelation of similar samples during classification, we strive to ensure the integrity of the plots when drawing polygons and pay special attention to collecting non-coffee plots around the coffee plots when collecting coffee polygons to increase the representativeness of the samples. Meanwhile, we used the 2bulu app (<https://www.2bulu.com/>) to take photographic records, with over half of the plots having one close-up and one distant photo each, while the majority of the remaining plots have at least one on-site image, providing robust visual verification for ground-truthing (Table 3). Our sample covers all coffee-growing towns in the study area. For areas that cannot be reached in person, we use high-resolution images from Google Earth to assist in sample collection. Because the goal of our model is only to distinguish between coffee and non-coffee, this strategy can maximize the model's ability to distinguish between coffee and non-coffee.

Due to the large area of the polygon, using all 10 m grid data for modeling would require too much computation, so we need to perform sampling. At the same time, our polygon area varies, and larger polygons generate more sample points during sampling, which may lead to unbalanced samples.

Table 2. Information of different land cover samples and model dataset.

Land Cover Type	Binary Classification Scheme	Polygon Numbers	Total Area (km ²)	Training Points	Testing Points
Coffee	Coffee	938	14.44	13,613	5910
Cropland	Non-coffee	816	15.89	12,641	3286
Sparse or low veg	Non-coffee	734	20.47	16,780	4509
Dense veg	Non-coffee	1028	120.8	36,574	10,504
Bare land	Non-coffee	91	4.56	1930	787
Built-up	Non-coffee	44	13.0	1795	465
Water	Non-coffee	44	13.97	1711	494
Total		3695	203.13	85,044	25,955

Table 3. Land cover type with example high-resolution imagery from Google Earth imagery and field photo.

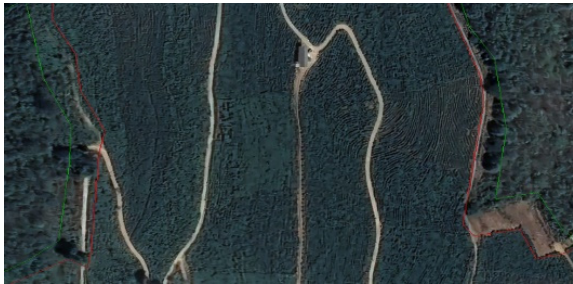
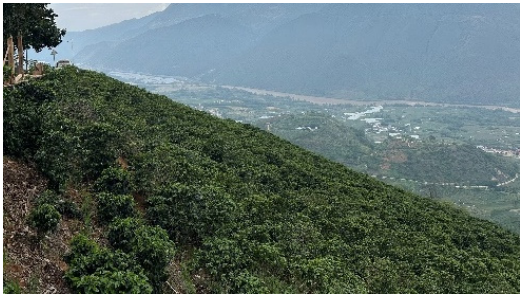
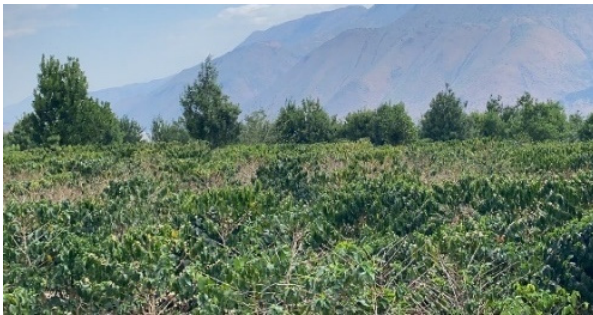
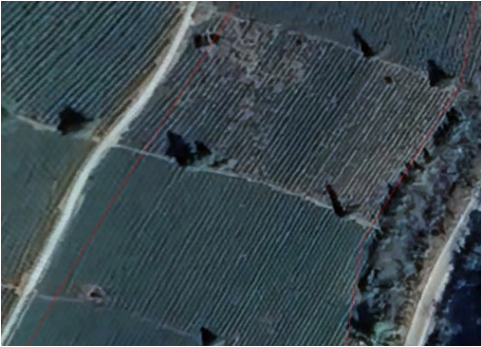
Land Cover	Image	Photo
Pure coffee plot		
Coffee plot with a small amount of nut shading trees		
Rubber-covered coffee plantation with full shade (Red polygon); Pure rubber (green polygon)		

Table 3. Cont.

Land Cover	Image	Photo
Rubber trees		
Nuts		
Mango field		
Sugarcane field		

Table 3. Cont.

Land Cover	Image	Photo
Tea		

To avoid over sampling in large polygons, we set a maximum sample size by using a threshold of polygon area. It yields a trade-off: when the threshold is low, there would be more sample points generated from scattered small polygons, which may lead to a certain degree of noise, making it difficult for the model to learn the classification features. If the threshold is too high, polygons with larger areas generate an excessive number of sample points, causing the features learned by the model to be concentrated in the corresponding regions, thereby reducing the model’s generalization ability. To decide the potentially best area threshold, we tested thresholds of 1, 2, and 3 hectares. By observing the accuracy in classification results (Table S1), we ultimately determined the cutoff threshold to be two hectares.

We also tested the impact of different sampling densities on model accuracy, including one point per 1000 m², 500 m², and 300 m². As the sample density increases, there is a slight improvement in model accuracy (Table S2), possibly because more training samples provide a more balanced class distribution, allowing the classifier to better learn features from different classes and thus improve accuracy. Finally, we chose the sample point with the highest accuracy of every 300 square meters.

When dividing the training and testing sets, a 1 km grid was first generated to cover the entire study area. Each sample polygon was assigned to a corresponding grid cell based on its geometric centre position, and the polygon areas within each grid cell were calculated. To ensure a balanced distribution of sample categories, the grids were then randomly divided into training and testing sets according to a ratio of 80% to 20% of the total polygon area. At the same time, all sample points generated from the same polygon are either used for training or for testing. Additionally, to maintain the representativeness of different land cover types, the area ratio of polygons for each category between the training and testing sets was controlled within a range of 4:1 to 6:4. Figure 3 shows the position of the geometric center of the polygon used for training and testing.

2.3.3. Census Data

To further validate the accuracy of the generated coffee distribution map for 2023, agricultural census data was collected from the statistical yearbooks of each county. This data was accessed through the National Bureau of Statistics and the Yunnan Provincial Bureau of Statistics website (<http://www.stats.gov.cn/>, last accessed on 8 July 2024). The census data provides independent, authoritative reference information for cross-validation and supports the reliability assessment of the mapping results.

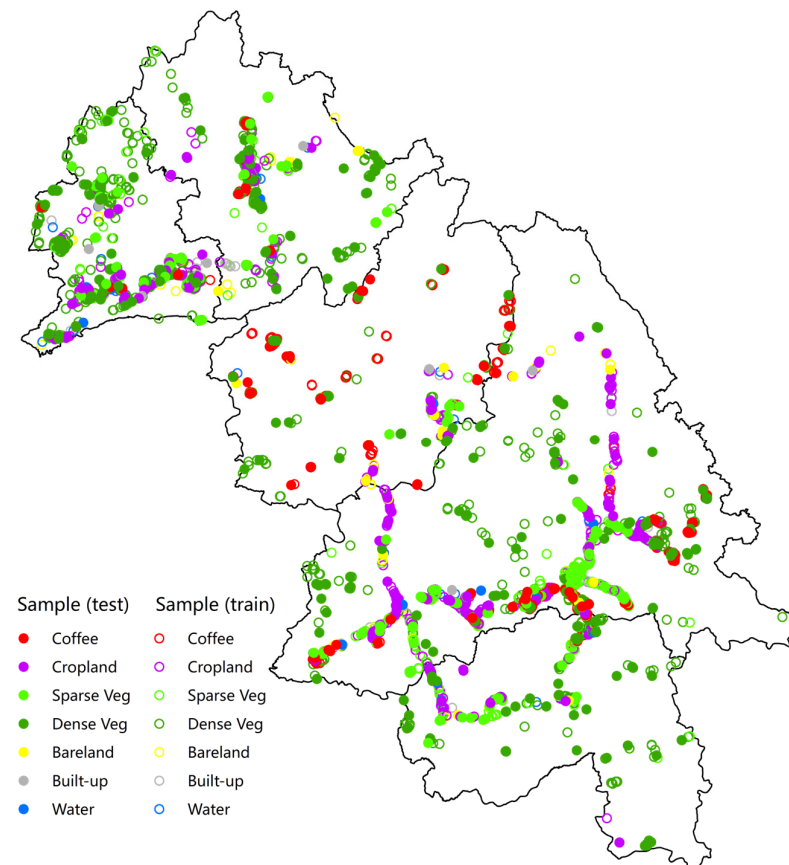


Figure 3. Sample distribution.

2.4. Method

2.4.1. SNIC Segmentation

Simple Non-Iterative Clustering (SNIC) is an advanced super-pixel segmentation algorithm derived from Simple Linear Iterative Clustering (SLIC). It offers advantages such as lower memory consumption and faster processing speed, demonstrating significant potential in land use and land cover classification, including crop mapping [38]. The SNIC algorithm initiates the process by placing clustering seeds on a regular grid and assigns pixels to super-pixel clusters based on their spatial and spectral distance from the seeds. Subsequently, a priority queue is employed to iteratively cluster and update the centroid positions, ensuring an efficient and adaptive segmentation process.

In this study, we applied the SNIC algorithm to the various spectral bands of the five Sentinel-2 composite images, resulting in a total of 50 bands. To identify the optimal parameters for accurately delineating coffee plots, we systematically adjusted the key parameter of super-pixel size from 4 to 48 in increments of 4. Super-pixel size determines the spatial interval between cluster seeds, directly affecting the granularity of the generated segments. We use visual inspection together with the Intersection over Union (IoU) metrics to decide the super-pixel size. As shown in Figure 4 below, when the segmentation sizes were 4, 8, 12, and 16, the average IoU of all polygons were 0.27, 0.43, 0.34, and 0.28, respectively. Correspondingly, we take the super-pixel size (8) with the largest IoU value (0.43) in subsequent analyses, which provided the best balance between preserving the integrity of small, fragmented coffee plots and avoiding the over-merging of distinct land cover categories. This parameter effectively mitigated the fragmentation caused by excessively small super-pixels while preventing the blending of heterogeneous regions associated with larger super-pixels.

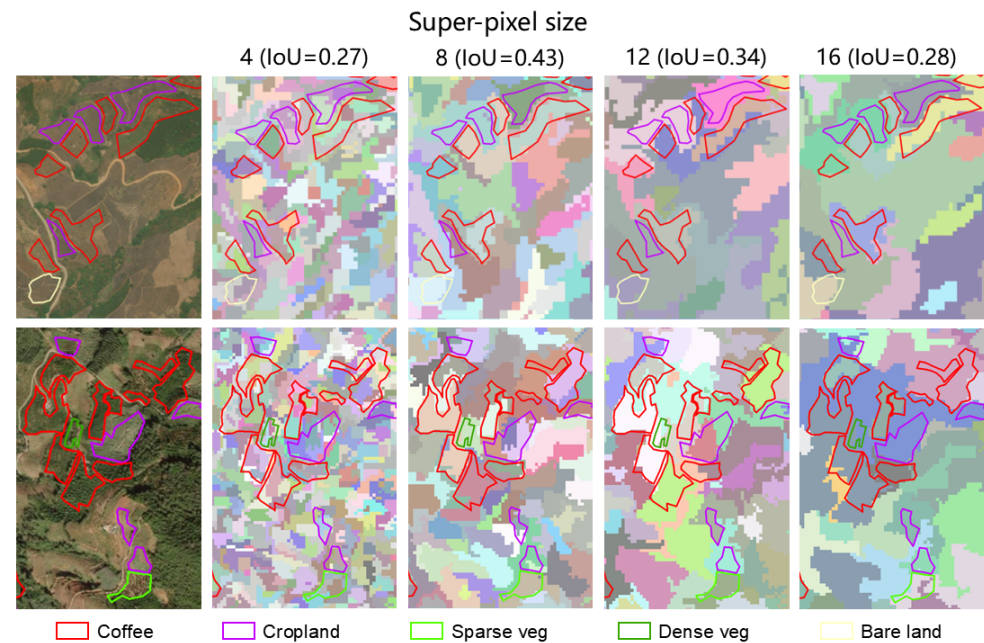


Figure 4. Comparison of different super-pixel size.

For the segmented images obtained, we extract training and testing data from the sample point data to build a classification model, and use the best-performing model to classify the segmented images. Since the values of pixels in the same band within the same object become the same after segmentation, we are actually classifying the image at the object level.

2.4.2. Deep Learning Classification

To predict the distribution of coffee, we employed a pixel-based universal deep neural network (DNN) architecture (Figure 5). The median composites of Sentinel-2 images from five different time periods were processed through a pre-processing network to extract features conducive to coffee classification during each specific period. Specifically, the ten spectral bands of each median composite image were input into the corresponding ten nodes of the input layer. These inputs were then processed through two hidden layers with 12 nodes each, followed by an output layer with 4 nodes to extract the spectral features for each time period. The features derived from optical images were subsequently concatenated with terrain features and passed through a final prediction network. This prediction network consisted of three hidden layers with 16 nodes, 8 nodes and 4 nodes each, ultimately producing the classification results.

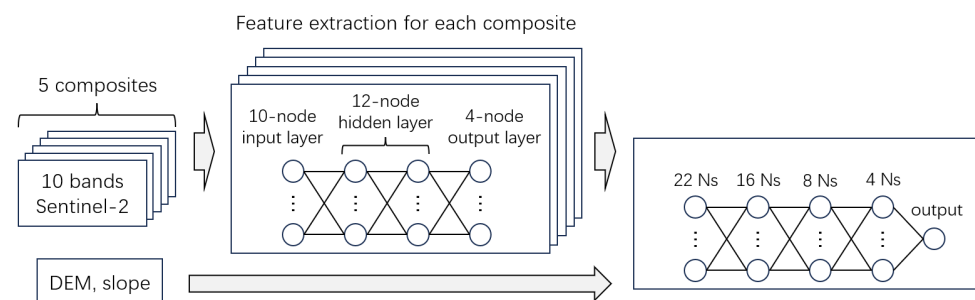


Figure 5. Deep Learning Network Architecture.

We explored different network structures and observed that directly inputting band data and terrain features from all five time periods into a fully connected network reduced

classification accuracy under the same parameter size. This suggests that the pre-processing network effectively extracts time-specific features, enhancing model performance.

During the training phase, we adopted different loss functions based on the classification task. For binary classification problems, we used the Binary Cross Entropy loss function, while for multi-class classification tasks, we applied the Cross-Entropy loss function. The Adam optimizer was chosen for model optimization, as it combines the advantages of both Momentum and RMSprop optimizers (Table 4). Specifically, Adam calculates both the first-order moment estimate (momentum) and the second-order moment estimate of the gradient, which leads to more stable and efficient convergence. We used the default learning rate and weight decay settings provided by the PyTorch (v2.10.0) library. The model was trained for a total of 200 epochs, which we found to be sufficient for the evaluation metrics on the test set to reach a stable level.

Table 4. List of hyperparameters of DNN model.

Hyperparameter	Value
Learning rate	0.001
Eps	1×10^{-8}
Weight decay	0

After obtaining the training data, we used the PyTorch library in Python (v3.10.1) to train the network parameters. Once the model was trained, we deployed the parameters to a cloud platform to predict coffee categories using the Python API provided by GEE and geemap (v0.36.1). Geemap is a Python package that facilitates interactive geospatial analysis and visualization within the GEE environment and allows users to analyze and visualize Earth Engine datasets interactively in a Jupyter v3-based environment. To further evaluate the performance of the deep neural network (DNN), we also conducted a random forest (RF) classification as a baseline comparison. The input data used for the DNN model consists of 5 time periods of raw image bands and 2 terrain features, for a total of 52 input features. Considering that machine learning models represented by random forests do not have feature mining capabilities, we added commonly used vegetation indices relevant to crop and coffee classification to the original input (Table 5), including the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Green Chlorophyll Vegetation Index (GCVI), Normalized Difference Water Index (NDWI), and Normalized Difference Tillage Index (NDTI).

Table 5. List of spectral indices used for random forest classification.

Name	Formula
NDVI	$\frac{nir-red}{nir+red}$
EVI	$\frac{2.5 \times (nir-red)}{((nir+6 \times red-7.5 \times blue)+1)}$
GCVI	$\frac{nir}{green} - 1$
NDWI	$\frac{green-nir}{green+nir}$
NDTI	$\frac{swir1-swir2}{swir1+swir2}$

2.4.3. Accuracy Assessment

To quantitatively evaluate the accuracy of our coffee distribution map, we used county-level census data and ground truth samples. The confusion matrix, based on the testing dataset, was employed to calculate key evaluation metrics (Table 6). These metrics include overall accuracy, producer accuracy, user accuracy, and F1 score, which assess the model's performance in classifying coffee distribution at various levels of certainty.

Table 6. Evaluation metrics.

Metric	Description
True Positive (TP)	Correctly extracted coffee sample points
False Positive (FP)	Background misclassified as coffee
True Negative (TN)	Sample points correctly categorized as background
False Negative (FN)	Coffee points misclassified as background
	$(TP + TN)/(TP + FP + TN + FN)$
Overall Accuracy (OA)	Proportion of correctly recognized coffee and background
User Accuracy/Precision (UA)	$TP/(TP + FP)$
	Proportion of real coffee in the results
Producer Accuracy/Recall (PA)	$TP/(TP + FN)$
	Proportion of extracted Coffee in validation data
F1 Score	$2 \times TP/(2 \times TP + FP + FN)$
	Weighted average of the precision and the recall

Additionally, the coefficient of determination (R^2) was computed to assess the correlation between the mapped coffee areas and the census data. The R^2 value is calculated using the following equation:

$$R^2 = 1 - \frac{\sum_{i=1}^n (s_i - y_i)^2}{\sum_{i=1}^n (s_i - \bar{s})^2} \quad (1)$$

where s_i and y_i represents the statistical (census-based) and mapped coffee area for the county i , $\bar{s}$ is the average statistical area, and n represents the total number of counties.

Finally, confusion matrices were generated based on the predicted coffee areas and the ground truth data to evaluate the classification performance and identify any potential misclassifications between coffee and non-coffee areas. These matrices were instrumental in providing a detailed breakdown of classification errors.

2.4.4. Feature Importance Analysis Using SHAP

To interpret the decision-making process of the deep neural network and to identify the key features contributing to coffee distribution mapping, we employed the SHapley Additive exPlanations (SHAP) framework. SHAP is a game-theoretic approach that quantifies the contribution of each input feature to the model prediction by attributing the difference between the model output and its expected value to individual features.

In this study, SHAP analysis was conducted at both global and feature-response levels. First, a SHAP bar plot was generated to evaluate the global importance of all input features. The importance of each feature was quantified as the mean absolute SHAP value across all test samples, which reflects its overall contribution to the model predictions regardless of direction. This global ranking allowed us to identify the most influential spectral bands and vegetation indices derived from dry- and wet-season Sentinel-2 composite images.

Based on the global SHAP importance ranking, the top four most important features were selected for further analysis. For each of these features, SHAP dependence plots were generated to examine the relationship between feature values and their corresponding SHAP values. The dependence plots illustrate how variations in a specific feature influence the model output while implicitly accounting for interactions with other features. This analysis provides insights into the nonlinear and potentially non-monotonic responses learned by the deep neural network.

By combining the SHAP bar plot and the SHAP dependence plots, the proposed approach enables a comprehensive interpretation of the model behavior, including both the relative importance of input features and their detailed response patterns. This interpretability

analysis helps to reveal the spectral and seasonal characteristics that are most relevant for distinguishing coffee plantations from other land cover types in Yunnan Province.

3. Results

3.1. Coffee Distribution Map

To enhance the visual representation of the coffee distribution, we resampled the 10 m-resolution coffee distribution data to a 1 km grid (Figure 6), which offers a more intuitive visual understanding of the coffee distribution across the region. The results reveal that coffee cultivation is not uniformly distributed across the study area but is highly concentrated in the dry hot valley regions.

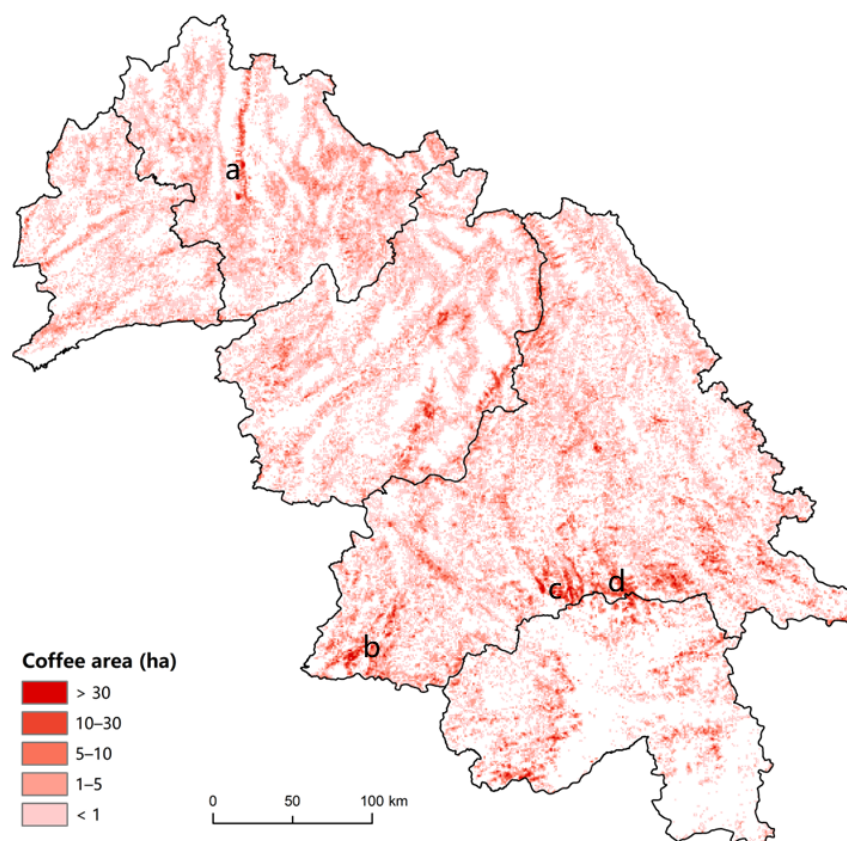


Figure 6. Map of coffee distribution derived from object-based DNN classification in 2023 (aggregate from 10 m grids to 1 km grids). ‘a’–‘d’ corresponds to the verification locations in figures below.

3.2. Validation

The generated coffee maps were evaluated from three aspects: (1) the overall classification accuracy on the testing data (Table 7), (2) the consistency between the mapped coffee planting area and the areas recorded in the statistical yearbooks (Figure 7), and (3) the comparison between the mapping results and ground control polygons using high-resolution images (Figures 8 and 9).

Table 7. Comparison of classification accuracy under different classification methods and number of categories.

Method	Categories	OA	UA	PA	F1
RF	2	0.81	0.84	0.47	0.6
RF	7	0.79	0.72	0.58	0.65
DNN	2	0.87	0.90	0.96	0.93
DNN	7	0.86	0.91	0.93	0.92

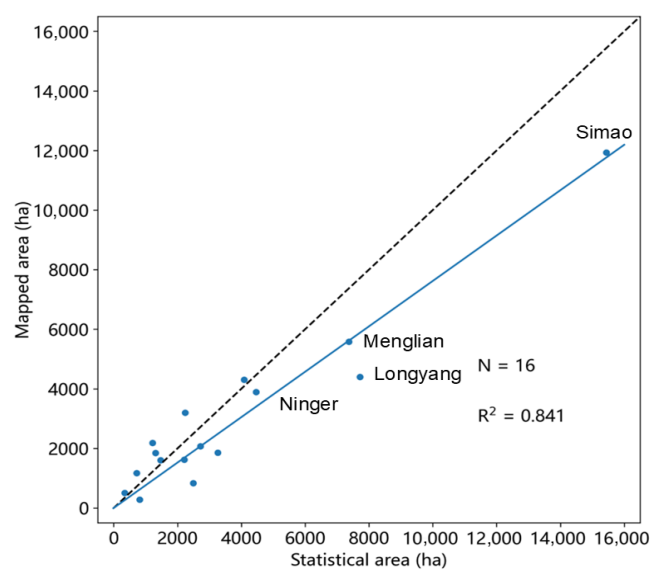


Figure 7. Comparison of coffee areas with statistics of 2023 at the county level (solid and dashed lines represent 1:1 reference line).

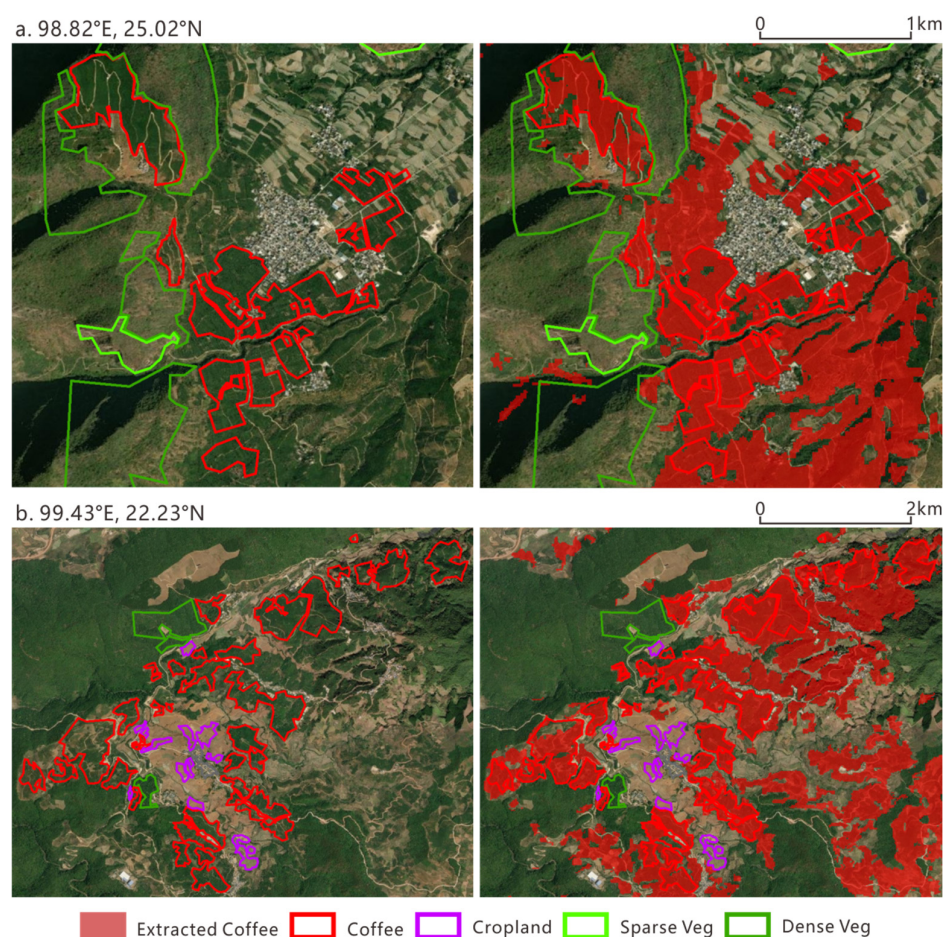


Figure 8. Details of coffee distribution compared with high-resolution image in site (a,b) shown in Figure 6 (left: high-resolution image with sample polygon; right: extracted coffee distribution).

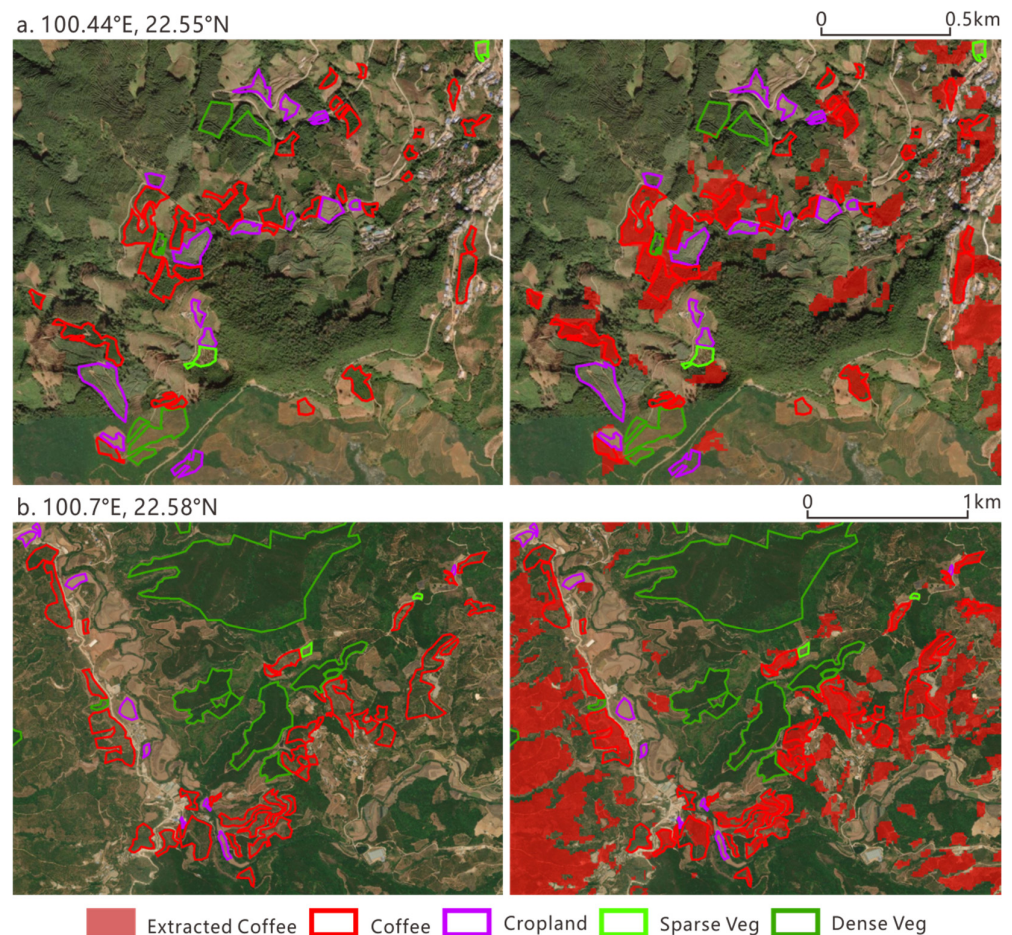


Figure 9. Details of coffee distribution compared with high-resolution image. (a) site c shown in Figure 6. (b) site d shown in Figure 6. (left: high-resolution image with sample polygon; right: extracted coffee distribution).

Using SNIC segmentation with super-pixel size of 8 and DNN, the model achieved overall accuracy (OA) of 0.87, user accuracy (UA) of 0.90, producer accuracy (PA) of 0.96, an F1 score of 0.93, and an AUC of 0.96 (Figure S3). These results indicate a high level of classification performance. Compared to random forest classification, the DNN model significantly improved accuracy. Random forest, on the other hand, is more sensitive to the number of classification categories. As the number of categories decreases, the accuracy of the classification tends to decrease when multiple non-coffee categories are merged into one. The confusion matrix of the multi class model is shown in Tables S1 and S2. In contrast, the DNN model is less affected by the number of categories, offering more stable classification accuracy.

To assess the accuracy of our coffee distribution map, we compared the coffee-planting areas derived from our annual coffee maps with the records found in the 2023 statistical yearbooks. As shown in Figure 7, the coefficient of determination (R^2) is 0.841 in 2023, which demonstrates a reasonable level of consistency between our mapped data and the statistical records. However, when comparing the mapped coffee planting area with the figures in the statistical yearbook, we found that in most counties, the mapped areas were smaller than those recorded in the yearbook. This suggests that some coffee plots were not identified in the mapping process. There are several possible reasons for this discrepancy: first, some coffee plots are small and scattered, with spectral characteristics that closely resemble surrounding vegetation, making them challenging to distinguish. Second, in certain plots, the density of natural shade trees is high, which can obscure the spectral

signals of the coffee plants. This issue is more prevalent in remote northern areas where coffee cultivation has a longer history and where planting techniques are relatively less advanced. Lastly, some plots have lower coffee planting densities or involve intercropping with other crops, reducing the clarity of the coffee's spectral signature (Figure S4), especially in southern regions with higher coffee cultivation, such as part of the unextracted coffee in the northwest corner plot in Figure 8a.

3.3. Feature Importance Analysis Result

The global SHAP feature importance analysis reveals that the most influential variables for coffee distribution mapping in Yunnan Province are dominated by Sentinel-2 shortwave infrared (SWIR) and red-edge bands, with clear seasonal distinctions. The top eight features, ranked by the mean absolute SHAP values, are *swir1_Nov*, *re1_Feb*, *swir1_Feb*, *re2_AprToOct*, *swir2_Nov*, *re2_Nov*, *swir2_Feb*, and *re3_Nov* (Figure 10).

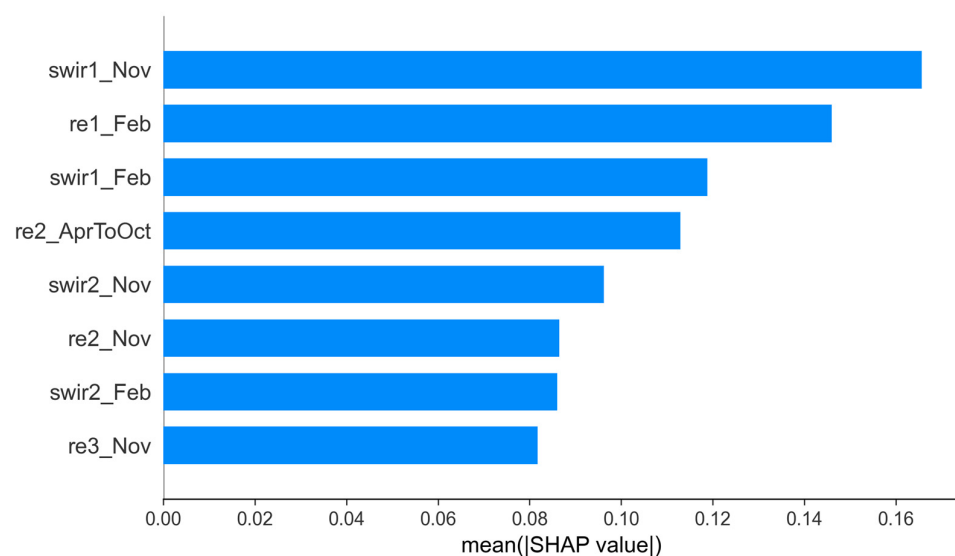


Figure 10. Global importance of the top eight important features. ‘AprToOct’ represents the composite image from April to October, ‘Nov’ represents the composite image of November, and ‘Feb’ represents the composite image of February.

Among these features, SWIR bands (SWIR1 and SWIR2) account for five out of the eight most important variables, indicating that moisture-related spectral information plays a dominant role in the discrimination of coffee plantations. In particular, *swir1_Nov* (November composite) and *swir1_Feb* (February composite) exhibit the highest contributions, highlighting the importance of dry-season canopy and soil moisture conditions for coffee identification.

Red-edge features also show strong importance across different seasonal composites. The presence of *re1_Feb*, *re2_AprToOct*, *re2_Nov*, and *re3_Nov* among the top-ranked features suggests that vegetation structural and chlorophyll-related information is critical for distinguishing coffee plantations from other land cover types. Notably, red-edge features derived from both the wet season (April–October) and dry season (November and February) contribute substantially to the model predictions.

Seasonal analysis of the selected features indicates that dry-season composites (November and February) contribute more prominently to the global feature importance than the wet-season composite. This pattern implies that spectral differences between coffee plantations and surrounding vegetation are more pronounced during the dry season, when phenological and moisture contrasts are enhanced.

SHAP dependence plots for the top eight features further illustrate their nonlinear relationships with model predictions. Several SWIR-related features exhibit non-monotonic response patterns, reflecting the heterogeneous moisture conditions and management practices of coffee plantations in Yunnan Province. Red-edge features show clear sensitivity to variations in vegetation structure and physiological status, reinforcing their role in coffee discrimination.

The SHAP dependence plot for *swir1_Nov* exhibits a clear U-shaped pattern (Figure 11), indicating that both low and high SWIR1 values in November contribute positively to the model's coffee predictions, whereas intermediate values tend to have lower SHAP contributions. This non-monotonic relationship suggests that coffee plantations are associated with heterogeneous moisture-related spectral responses during the early dry season.

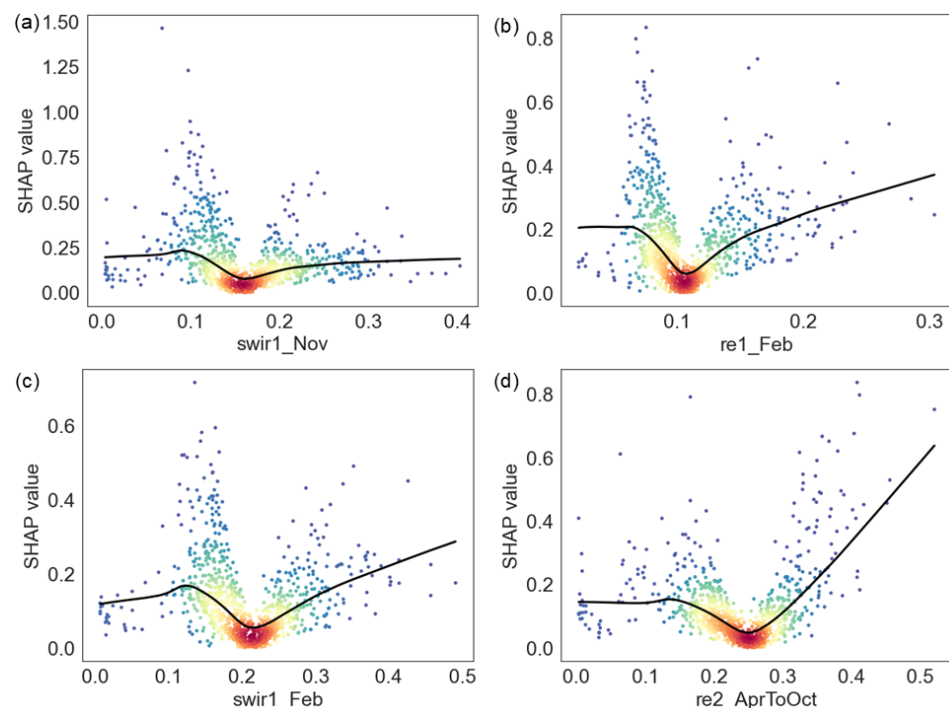


Figure 11. SHAP dependence plot for the top 4 important features. (a) *swir1* in November; (b) *re1* in February; (c) *swir1* in February; and (d) *re2* from April to October.

The dependence plot of *re1_Feb* also shows a pronounced U-shaped response, with positive SHAP values observed at both lower and higher red-edge reflectance levels in February. This pattern indicates that variations in vegetation structure or chlorophyll-related properties under different canopy and management conditions can both support coffee discrimination during the late dry season.

For *swir1_Feb*, a similar U-shaped dependence is observed, with increased SHAP values at both extremes of SWIR1 reflectance in February. Compared with the November composite, the February SWIR1 feature shows a wider range of high SHAP values, suggesting enhanced spectral separability of coffee plantations under more pronounced dry-season conditions.

The SHAP dependence plot of *re2_AprToOct* derived from the wet-season composite (April–October) reveals a U-shaped relationship as well. Although overall SHAP magnitudes are lower than those of dry-season features, both low and high red-edge values still contribute positively to coffee predictions, indicating that structural variability of coffee canopies persists during the wet season.

4. Discussion

4.1. Potential Values of the Coffee Distribution Map in Yunnan, China

By integrating Sentinel-2 multi-spectral imagery with DEM-derived topographic data, this study has produced the first high-precision coffee distribution map for the entirety of Yunnan Province in 2023. This achievement fills a critical gap in spatially refined data for China's coffee-growing regions and provides an essential academic supplement to the global coffee distribution dataset, enabling international researchers to objectively examine the expansion of coffee into the tropical and subtropical fringes of East Asia.

The scientific value of this dataset extends across multiple disciplines, particularly in agricultural monitoring, climate change, and landscape ecology. Researchers can leverage this high-resolution map to track the regional expansion of cultivation [39–42], investigate how shifting climate envelopes influence planting suitability [43], and analyze the environmental signatures of different cultivation zones [44]. Furthermore, given the strong correlation between coffee expansion and environmental concerns—such as potential deforestation, land degradation, and biodiversity loss in sensitive watersheds—this dataset provides a spatial baseline for ecologists to explore the interactions between coffee agroecosystems and surrounding natural habitats [45]. While the current map does not yet differentiate between various planting densities or shading methods, it provides a “base map” that can be refined with supplementary high-resolution data for nuanced analyses of complex agroforestry patterns.

At the practical level, the spatial information provided here is foundational for the transformation and upgrading of Yunnan's coffee industry. As the region shifts toward a “premium bean” strategy, accurate distribution data becomes indispensable for implementing standardized management, assessing climate risks, and securing geographical indication (GI) protections. Furthermore, this dataset supports sustainable development and agricultural policy research. Government and non-governmental organizations can utilize these findings to formulate more rational cultivation strategies, optimize resource allocation, and enhance productivity through rigorous monitoring of land-use changes. Ultimately, this work marks a transition from traditional statistical estimation to precise, spatially explicit management for the Chinese coffee industry, aligning it with international standards of supply chain transparency and environmental accountability [46].

4.2. Advantages of Object-Based Deep Learning for Coffee Mapping

Our study demonstrates that the integration of object-based image segmentation and deep learning provides an effective framework for coffee plantation mapping in Yunnan Province. By applying the Simple Non-Iterative Clustering (SNIC) algorithm to Sentinel-2 composite images prior to model training, spatially coherent image objects were generated, which better represent the structural characteristics of coffee plantations compared to merely pixel-based approaches. Coffee plantations in Yunnan are typically characterized by irregular shapes, mixed canopy structures, and varying degrees of shading, which can be more effectively captured at the object level.

Our framework—based on Sentinel-2 imagery, SRTM terrain features, SNIC segmentation, and a deep neural network (DNN)—achieved UA, PA, and F1-scores of 0.90, 0.96, and 0.93, respectively, for the coffee class across the main production areas of Yunnan Province. Notably, the producer's accuracy (0.96) and F1-score (0.93) exceed those reported in previous studies, despite relying on fewer data sources. For instance, Maskell mapped coffee in Dak Lak, Vietnam, using Sentinel-1, Sentinel-2, and SRTM-derived terrain features, achieving a user's accuracy (UA), producer's accuracy (PA), and F1-score of 0.89, 0.66, and 0.76 for the coffee class [10]. Escobar-López integrated Sentinel-1, Sentinel-2, ALOS PALSAR, SRTM elevation, and climatic variables to map coffee in the Sierra Madre de

Chiapas, Mexico, reporting UA, PA, and F1-scores of 0.92, 0.89, and 0.90 [9]. Similarly, Tridawati combined pan-sharpened GeoEye-1, multi-temporal Sentinel-2, and DEM data to map coffee in Mt. Puntang, Indonesia, achieving UA, PA, and F1-scores of 0.90, 0.92, and 0.91 [11].

Two methodological factors likely contributed to this performance gain. First, object-based segmentation using the SNIC algorithm aggregates spectrally homogeneous pixels into spatially coherent units prior to classification. Object-based image analysis has been widely recognized as advantageous for heterogeneous vegetation mapping because it reduces within-class spectral noise and better captures structural patterns [17,47,48]. In mountainous coffee-growing regions, plantations often exhibit irregular boundaries, mixed canopy structures, and terrain-induced illumination differences. By shifting the classification unit from pixel to object, the model better represents plantation-scale characteristics rather than isolated spectral responses, thereby mitigating the salt-and-pepper effect commonly observed in pixel-based RF classifications.

Second, the DNN classifier enhances the capacity to model nonlinear feature interactions among multi-seasonal spectral bands and topographic variables. While RF is robust and widely used in crop mapping [49], it relies on ensemble decision trees that partition the feature space through axis-aligned splits. In contrast, deep neural networks can learn high-dimensional nonlinear relationships and hierarchical representations, which are particularly relevant for perennial crops such as coffee that exhibit substantial intra-class variability due to management practices, canopy density differences, elevation gradients, and seasonal phenology [19,50,51]. The higher PA and F1-score obtained in this study suggest that the object-based DNN framework is better suited to capturing these complex and heterogeneous spectral patterns.

Overall, the results indicate that integrating object-based segmentation with deep learning provides measurable advantages over traditional pixel-based RF approaches in coffee mapping, even without incorporating radar or climate datasets. This framework offers a scalable and data-efficient solution for high-resolution perennial crop mapping in topographically complex environments.

4.3. Important Feature in Coffee Mapping According to SHAP Analysis

SHAP-based feature importance analysis offered deeper insights into the biophysical drivers of coffee distribution in Yunnan. The most influential variables in our model were November SWIR1, February red-edge band 1 (re1), February SWIR1, and wet-season (April–October) red-edge band 2 (re2). These findings underscore the critical role of both shortwave infrared (SWIR) and red-edge spectral regions, particularly during the dry-to-early growing season transition, in accurately mapping Arabica coffee plants.

Our findings further reveal that dry-season canopy dynamics are central to coffee mapping, as evidenced by the strong influence of November and February SWIR1. SWIR bands are highly sensitive to leaf water content and internal canopy structure, making them effective for detecting differences in perennial crop vigor and management intensity [52]. In subtropical mountainous regions such as Yunnan, coffee plants often experience pronounced dry-season water stress, which can enhance spectral separability from surrounding evergreen forests or other crops. This interpretation is supported by research in Mexico [9], where SWIR reflectance (B11) during the dry-to-wet transition (April–May) proved critical for coffee classification, highlighting the relevance of moisture-sensitive wavelengths during transitional phenological stages.

The importance of red-edge bands (re1 and re2) further underscores the role of chlorophyll content and canopy physiological condition in coffee mapping. Sentinel-2 red-edge bands are widely recognized for their sensitivity to subtle variations in vegetation health

and bio-mass [53]. This is consistent with work in Mexico that January NIR (B8, B8A) and red-edge bands (B5, B7) were dominant predictors in the Sierra Madre de Chiapas [9], indicating that chlorophyll-related signals are consistently important for coffee mapping across regions. Our SHAP results not only align with these findings but also extend them by showing that red-edge importance extends into the wet season (April–October). This suggests that sustained canopy vigor differences during peak growth periods contribute to improved separability in Yunnan.

In contrast, studies in Vietnam and Indonesia emphasize somewhat different feature importance patterns, underscoring the role of geographic context. In Dak Lak, dry-season NDTI, wet-season red-edge (re2), radar texture (GLCM IDM from descending VH), and elevation were identified as key variables, highlighting the contribution of soil exposure and canopy texture information [10]. In Mt. Puntang, Indonesia, that elevation and slope were reported to be dominant predictors, followed by dry-season humidity and visible bands (blue and green) [11]. These results suggest that topographic gradients and microclimatic controls can strongly shape coffee distribution in certain mountainous landscapes. While elevation did not rank among the top SHAP features in our Yunnan model, terrain information was included and likely contributed indirectly through interactions with spectral variables. The relatively lower standalone importance of topographic features in Yunnan may reflect the broader spatial extent and more heterogeneous plantation conditions, where spectral–phenological signals dominate over purely topographic constraints.

4.4. Limitation

Although our coffee mapping model demonstrates high accuracy compared to existing remote sensing-based coffee mapping models, there are still some uncertainties, which are reflected in the consistent estimation of coffee areas compared to census data.

The main sources of the uncertainties are the low quality and low availability of remote sensing data, as well as the feature mining capabilities of the coffee mapping model. Due to the limitation of remote sensing image resolution and contamination in cloudy or rainy days, our model exhibits certain limitations in identifying coffee crops in areas with unclear land boundaries or mixed planting, as well as in fragmented or obscured coffee plots (with a plot width of less than 10 m). In the future, higher resolution data sources such as PlanetScope can be combined. Meanwhile, fully connected neural networks have not been optimized for specific data structures. In the future, Transformer-based networks can be introduced to uniformly process time series of images of different lengths in tropical regions, in order to efficiently and comprehensively mine effective information. In part, the discrepancy may stem from the limited availability and low quality of statistical data on economic crops in underdeveloped mountainous regions. These data are often spatially coarse, outdated, inconsistently collected, and lack transparency. Without remote sensing or geospatial calibration, official estimates may misrepresent actual planting areas. Thus, our mapping results should complement, rather than replace, statistical data to better support policy-making and agricultural management.

5. Conclusions

This study produced the first 10 m resolution, regional-scale coffee plantation distribution dataset for Yunnan Province, China, for the year 2023, based on Sentinel-2 imagery and SRTM topographic data within the Google Earth Engine platform. By implementing an object-based mapping workflow and supervised classification, we generated a spatially explicit and internally validated coffee map with an overall accuracy of 0.87 and an F1 score of 0.93 for the coffee class. These results indicate that the dataset is sufficiently reliable for

regional-scale agricultural and environmental applications in complex mountainous landscapes.

Feature contribution analysis suggests that shortwave infrared and red-edge information, particularly during the dry season, plays an important role in coffee discrimination. This finding enhances confidence in the ecological relevance and stability of the mapping results and provides useful guidance for future perennial crop monitoring efforts.

Beyond the 2023 snapshot, the workflow established in this study is designed to be operational and reproducible, enabling future updates and the development of multi-annual time series products. The resulting high-resolution coffee distribution dataset provides a foundational spatial baseline for monitoring plantation expansion, assessing land-use and ecological impacts, and supporting sustainable coffee development in Yunnan. We expect that this dataset and workflow will serve as practical tools for researchers, land managers, policymakers, and industry stakeholders engaged in agricultural planning and environmental management in southwestern China.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/rs18060940/s1>, Table S1: Classification accuracy under different polygon sample clipping thresholds; Table S2: Classification accuracy under different sampling densities; Figure S1: County-level administrative divisions of the study area; Figure S2: Statistics on the number of available pixels in monthly remote sensing images of the research area; Figure S3: ROC curves of DNN classification model; Figure S4: Example of lower coffee planting densities inter-cropping with nuts (25°01′12″N, 98°49′47″).

Author Contributions: H.S. (Hongyu Shan) and T.Y. designed the study. W.Z. and X.C. provided technical suggestions. H.S. (Hao Sun) and Z.C. helped to collect ground sample. H.S. (Hongyu Shan) and T.Y. completed the model's training, the validation of the results, and wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study has been financially supported by the joint project of National Natural Science Foundation of China (NSFC, 72261147759) and Bill Melinda Gates Foundation (BMGF, 2022YFAG1004) and State Key Laboratory of Remote Sensing Science and Beijing Engineering Research Centre for Global Land Remote Sensing Products (OF202411).

Data Availability Statement: Data derived from the DNN binary classification were provided in Geotiff format and can be accessed on Zenodo (<https://zenodo.org/records/15031668>, accessed on 1 March 2025). This dataset represents the best coffee extraction results compared to other models and classification schemes used in the study. We provided two files. “2023_yunnan_coffee_10m.tif” is the 10 m coffee distribution data directly calculated based on remote sensing images provided by the GEE platform, which covers a spatial extent of 97.53°E to 102.32°E and 21.14°N to 25.86°N, encompassing 52,491 rows and 53,371 columns. A value of 1 indicates the presence of coffee at that location, while a value of 0 indicates the absence of coffee. “2023_yunnan_coffee_1km.tif” is the proportion of coffee area in a 1 km grid obtained by aggregating the previous 10 m coffee distribution data, encompassing 524 rows and 533 columns. The numerical range is 0–1, where 0 indicates that there is no coffee in the 1 km grid and 1 indicates that all coffee is present in the 1 km grid. The procedure of coffee mapping is conducted using Python and GEE. The code is available at <https://github.com/hyshan-geo/yunnan-cf-map-DL-GEE/tree/main> (accessed on 1 March 2025).

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

References

1. Peña-Barragán, J.M.; Ngugi, M.K.; Plant, R.E.; Six, J. Object-Based Crop Identification Using Multiple Vegetation Indices, Textural Features and Crop Phenology. *Remote Sens. Environ.* **2011**, *115*, 1301–1316. [CrossRef]
2. Dong, J.; Fu, Y.; Wang, J.; Tian, H.; Fu, S.; Niu, Z.; Han, W.; Zheng, Y.; Huang, J.; Yuan, W. Early-Season Mapping of Winter Wheat in China Based on Landsat and Sentinel Images. *Earth Syst. Sci. Data* **2020**, *12*, 3081–3095. [CrossRef]

3. Descals, A.; Wich, S.; Meijaard, E.; Gaveau, D.L.A.; Peedell, S.; Szantoi, Z. High-Resolution Global Map of Smallholder and Industrial Closed-Canopy Oil Palm Plantations. *Earth Syst. Sci. Data* **2021**, *13*, 1211–1231. [\[CrossRef\]](#)
4. Blickensdörfer, L.; Schwieder, M.; Pflugmacher, D.; Nendel, C.; Erasmi, S.; Hostert, P. Mapping of Crop Types and Crop Sequences with Combined Time Series of Sentinel-1, Sentinel-2 and Landsat 8 Data for Germany. *Remote Sens. Environ.* **2022**, *269*, 112831. [\[CrossRef\]](#)
5. Ponte, S. The ‘Latte Revolution’? Regulation, Markets and Consumption in the Global Coffee Chain. *World Dev.* **2002**, *30*, 1099–1122. [\[CrossRef\]](#)
6. Moat, J.; Williams, J.; Baena, S.; Wilkinson, T.; Gole, T.W.; Challa, Z.K.; Demissew, S.; Davis, A.P. Resilience Potential of the Ethiopian Coffee Sector under Climate Change. *Nat. Plants* **2017**, *3*, 17081. [\[CrossRef\]](#)
7. Hunt, D.A.; Tabor, K.; Hewson, J.H.; Wood, M.A.; Reymondin, L.; Koenig, K.; Schmitt-Harsh, M.; Follett, F. Review of Remote Sensing Methods to Map Coffee Production Systems. *Remote Sens.* **2020**, *12*, 2041. [\[CrossRef\]](#)
8. Pham, Y.; Reardon-Smith, K.; Mushtaq, S.; Cockfield, G. The Impact of Climate Change and Variability on Coffee Production: A Systematic Review. *Clim. Change* **2019**, *156*, 609–630. [\[CrossRef\]](#)
9. Escobar-López, A.; Castillo-Santiago, M.Á.; Hernández-Stefanoni, J.L.; Mas, J.F.; López-Martínez, J.O. Identifying Coffee Agroforestry System Types Using Multitemporal Sentinel-2 Data and Auxiliary Information. *Remote Sens.* **2022**, *14*, 3847. [\[CrossRef\]](#)
10. Maskell, G.; Chemura, A.; Nguyen, H.; Gornott, C.; Mondal, P. Integration of Sentinel Optical and Radar Data for Mapping Smallholder Coffee Production Systems in Vietnam. *Remote Sens. Environ.* **2021**, *266*, 112709. [\[CrossRef\]](#)
11. Tridawati, A.; Wikantika, K.; Susantoro, T.M.; Harto, A.B.; Darmawan, S.; Yayusman, L.F.; Ghazali, M.F. Mapping the Distribution of Coffee Plantations from Multi-Resolution, Multi-Temporal, and Multi-Sensor Data Using a Random Forest Algorithm. *Remote Sens.* **2020**, *12*, 3933. [\[CrossRef\]](#)
12. Gomez, C.; Mangeas, M.; Petit, M.; Corbane, C.; Hamon, P.; Hamon, S.; De Kochko, A.; Le Pierres, D.; Poncet, V.; Despinoy, M. Use of High-Resolution Satellite Imagery in an Integrated Model to Predict the Distribution of Shade Coffee Tree Hybrid Zones. *Remote Sens. Environ.* **2010**, *114*, 2731–2744. [\[CrossRef\]](#)
13. Kelley, L.C.; Pitcher, L.; Bacon, C. Using Google Earth Engine to Map Complex Shade-Grown Coffee Landscapes in Northern Nicaragua. *Remote Sens.* **2018**, *10*, 952. [\[CrossRef\]](#)
14. Cordero-Sancho, S.; Sader, S.A. Spectral Analysis and Classification Accuracy of Coffee Crops Using Landsat and a Topographic-environmental Model. *Int. J. Remote Sens.* **2007**, *28*, 1577–1593. [\[CrossRef\]](#)
15. Hebbar, R.; Ravishankar, H.M.; Trivedi, S.; Manjula, V.B.; Kumar, N.M.; Mukharib, D.S.; Mote, J.K.; Sudeesh, S.; Raj, U.; Raghuramulu, Y.; et al. National Level Inventory of Coffee Plantations Using High Resolution Satellite Data. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2019**, *XLII-3/W6*, 293–298. [\[CrossRef\]](#)
16. Yang, L.; Mansaray, L.R.; Huang, J.; Wang, L. Optimal Segmentation Scale Parameter, Feature Subset and Classification Algorithm for Geographic Object-Based Crop Recognition Using Multisource Satellite Imagery. *Remote Sens.* **2019**, *11*, 514. [\[CrossRef\]](#)
17. Yang, L.; Wang, L.; Abubakar, G.A.; Huang, J. High-Resolution Rice Mapping Based on SNIC Segmentation and Multi-Source Remote Sensing Images. *Remote Sens.* **2021**, *13*, 1148. [\[CrossRef\]](#)
18. Ma, L.; Li, M.; Ma, X.; Cheng, L.; Du, P.; Liu, Y. A Review of Supervised Object-Based Land-Cover Image Classification. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 277–293. [\[CrossRef\]](#)
19. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N. Prabhat Deep Learning and Process Understanding for Data-Driven Earth System Science. *Nature* **2019**, *566*, 195–204. [\[CrossRef\]](#)
20. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [\[CrossRef\]](#)
21. Ma, J.; Li, J.; He, H.; Jin, X.; Cesarino, I.; Zeng, W.; Li, Z. Characterization of Sensory Properties of Yunnan Coffee. *Curr. Res. Food Sci.* **2022**, *5*, 1205–1215. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Zhang, M.; Wang, R.; Li, Y.; Hu, X.; Li, M.; Zhang, M.; Duan, C. Ecological Suitability Zoning of Coffea Arabica L. in Yunnan Province. *Chin. J. Eco-Agric.* **2020**, *28*, 168–178. [\[CrossRef\]](#)
23. Zhang, S.; Liu, X.; Li, R.; Wang, X.; Cheng, J.; Yang, Q.; Kong, H. AHP-GIS and MaxEnt for Delineation of Potential Distribution of Arabica Coffee Plantation under Future Climate in Yunnan, China. *Ecol. Indic.* **2021**, *132*, 108339. [\[CrossRef\]](#)
24. Dong, W.; Hu, R.; Long, Y.; Li, H.; Zhang, Y.; Zhu, K.; Chu, Z. Comparative Evaluation of the Volatile Profiles and Taste Properties of Roasted Coffee Beans as Affected by Drying Method and Detected by Electronic Nose, Electronic Tongue, and HS-SPME-GC-MS. *Food Chem.* **2019**, *272*, 723–731. [\[CrossRef\]](#)
25. Dai, M.; Wu, L.; Xiang, X.; Zhang, Z.; Peng, Y. Risk Analysis of Meteorological Index Insurance for Coffee Chilling Injury in Yunnan. *South China Agric.* **2018**, *12*, 92–95. [\[CrossRef\]](#)
26. Li, M.; Dou, X.; Zhang, M.; Lu, W.; Zhou, J.; Zhu, Y. Evaluation of Drought Risk of Coffee Arabica in Yunnan Province. *Chin. J. Trop. Agric.* **2021**, *41*, 33–40.
27. Li, T. Natural Disasters and Integrated Disaster Risk Management in Yunnan. *West J.* **2023**, *17*, 10–14. [\[CrossRef\]](#)

28. Li, R.; Laroche, M.; Richard, M.-O.; Cui, X. More than a Mere Cup of Coffee: When Perceived Luxuriousness Triggers Chinese Customers' Perceptions of Quality and Self-Congruity. *J. Retail. Consum. Serv.* **2022**, *64*, 102759. [CrossRef]
29. Yunnan Provincial Department of Agriculture and Rural Affairs Climate and Resource Situation in Yunnan Province. Available online: https://nync.yn.gov.cn/html/2022/shushuoyunnansannong_0718/388503.html?cid=4977 (accessed on 15 March 2025).
30. Rigal, C.; Xu, J.; Hu, G.; Qiu, M.; Vaast, P. Coffee Production during the Transition Period from Monoculture to Agroforestry Systems in near Optimal Growing Conditions, in Yunnan Province. *Agric. Syst.* **2020**, *177*, 102696. [CrossRef]
31. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]
32. Defourny, P.; Bontemps, S.; Bellemans, N.; Cara, C.; Dedieu, G.; Guzzonato, E.; Hagolle, O.; Inglada, J.; Nicola, L.; Rabaute, T.; et al. Near Real-Time Agriculture Monitoring at National Scale at Parcel Resolution: Performance Assessment of the Sen2-Agri Automated System in Various Cropping Systems around the World. *Remote Sens. Environ.* **2019**, *221*, 551–568. [CrossRef]
33. Louis, J.; Pflug, B.; Main-Knorn, M.; Debaecker, V.; Mueller-Wilm, U.; Iannone, R.Q.; Giuseppe Cadau, E.; Boccia, V.; Gascon, F. Sentinel-2 Global Surface Reflectance Level-2a Product Generated with Sen2Cor. In *Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium*; IEEE: New York, NY, USA, 2019; pp. 8522–8525.
34. Main-Knorn, M.; Pflug, B.; Louis, J.; Debaecker, V.; Müller-Wilm, U.; Gascon, F. Sen2Cor for Sentinel-2. In *Proceedings of the Image and Signal Processing for Remote Sensing XXIII*; Bruzzone, L., Bovolo, F., Benediktsson, J.A., Eds.; SPIE: Warsaw, Poland, 2017; p. 3.
35. Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A | Earth Engine Data Catalog | Google for Developers. Available online: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED (accessed on 2 February 2025).
36. Housman, I.; Chastain, R.; Finco, M. An Evaluation of Forest Health Insect and Disease Survey Data and Satellite-Based Remote Sensing Forest Change Detection Methods: Case Studies in the United States. *Remote Sens.* **2018**, *10*, 1184. [CrossRef]
37. Schmitt, M.; Hughes, L.H.; Qiu, C.; Zhu, X.X. Aggregating Cloud-Free Sentinel-2 Images with Google Earth Engine. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2019**, *IV-2/W7*, 145–152. [CrossRef]
38. Mahdianpari, M.; Jafarzadeh, H.; Granger, J.E.; Mohammadimanesh, F.; Brisco, B.; Salehi, B.; Homayouni, S.; Weng, Q. A Large-Scale Change Monitoring of Wetlands Using Time Series Landsat Imagery on Google Earth Engine: A Case Study in Newfoundland. *GISci. Remote Sens.* **2020**, *57*, 1102–1124. [CrossRef]
39. Ortega-Huerta, M.A.; Komar, O.; Price, K.P.; Ventura, H.J. Mapping Coffee Plantations with Landsat Imagery: An Example from El Salvador. *Int. J. Remote Sens.* **2012**, *33*, 220–242. [CrossRef]
40. Mukashema, A.; Veldkamp, A.; Vrieling, A. Automated High Resolution Mapping of Coffee in Rwanda Using an Expert Bayesian Network. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *33*, 331–340. [CrossRef]
41. Le, Q.T.; Dang, K.B.; Giang, T.L.; Tong, T.H.A.; Nguyen, V.G.; Nguyen, T.D.L.; Yasir, M. Deep Learning Model Development for Detecting Coffee Tree Changes Based on Sentinel-2 Imagery in Vietnam. *IEEE Access* **2022**, *10*, 109097–109107. [CrossRef]
42. Chemura, A.; Mudereri, B.T.; Yalew, A.W.; Gornott, C. Climate Change and Specialty Coffee Potential in Ethiopia. *Sci. Rep.* **2021**, *11*, 8097. [CrossRef]
43. Senf, C.; Pflugmacher, D.; Van Der Linden, S.; Hostert, P. Mapping Rubber Plantations and Natural Forests in Xishuangbanna (Southwest China) Using Multi-Spectral Phenological Metrics from MODIS Time Series. *Remote Sens.* **2013**, *5*, 2795–2812. [CrossRef]
44. Bunn, C.; Läderach, P.; Ovalle Rivera, O.; Kirschke, D. A Bitter Cup: Climate Change Profile of Global Production of Arabica and Robusta Coffee. *Clim. Change* **2015**, *129*, 89–101. [CrossRef]
45. Jha, S.; Bacon, C.M.; Philpott, S.M.; Ernesto Méndez, V.; Läderach, P.; Rice, R.A. Shade Coffee: Update on a Disappearing Refuge for Biodiversity. *BioScience* **2014**, *64*, 416–428. [CrossRef]
46. Grabs, J. The Construction of Compliance with the European Union Deforestation Regulation in Global Coffee Value Chains. *Regul. Gov.* **2025**, 1–16. [CrossRef]
47. Blaschke, T. Object Based Image Analysis for Remote Sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]
48. Tassi, A.; Vizzari, M. Object-Oriented LULC Classification in Google Earth Engine Combining SNIC, GLCM, and Machine Learning Algorithms. *Remote Sens.* **2020**, *12*, 3776. [CrossRef]
49. Belgiu, M.; Drăguț, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
50. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
51. Bazame, H.C.; Molin, J.P.; Althoff, D.; Martello, M. Detection, Classification, and Mapping of Coffee Fruits during Harvest with Computer Vision. *Comput. Electron. Agric.* **2021**, *183*, 106066. [CrossRef]

52. Immitzer, M.; Vuolo, F.; Atzberger, C. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sens.* **2016**, *8*, 166. [[CrossRef](#)]
53. Delegido, J.; Verrelst, J.; Alonso, L.; Moreno, J. Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content. *Sensors* **2011**, *11*, 7063–7081. [[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.