Journal of Natural Resources and Environmental Management
http://dx.doi.org/10.29244/jpsl.15.6.1009

RESEARCH ARTICLE

OPEN ACCESS

Check for updates

Developing a Decision Tree Algorithm for Detecting Agroforestry and Monoculture Coffee Plantations Using Landsat 8 Imagery: A Case Study in Bandung Regency, Indonesia

Agasta Adhiguna, I Nengah Surati Jaya, Nining Puspaningsih

Department of Forest Management, Faculty of Forestry and Environment, IPB University, IPB Dramaga Campus, Bogor, 16680, Indonesia

Article History
Received 6 January 2025
Revised
22 September 2025
Accepted
26 September 2025

Keywords
brute force, coffee agroforestry, coffee monoculture, decision tree, machine learning

ABSTRACT

Data on the potential of coffee commodities in Bandung Regency is still mixed with data on other commodities. Therefore, the study aims to develop an algorithm that provides accurate spatial information through maps for both coffee plantations in agroforestry and monoculture systems. This study integrates the data derived from remotely sensed data and data derived using socio-geo-biophysical aspects, such as elevation, slope, distance from the road and rivers, proximity of the settlements, population density, proximity of villages, and a visually-based land-use-land cover map. The importance value for each variable was computed using several criteria, such as information gain, Gini index, and gain ratio. Meanwhile, the brute force method was applied to select the most significant variables in the model. The study found that the most significant variables for identifying coffee agroforestry and monoculture were ARVI, EVI, GARI, NRGI, and VDVI, as well as DEM, slope, proximity to roads, and visual-based LULC, using the criterion of information gain. The use of existing land-use and cover maps was the most influential variable in the model. The algorithm achieved an overall accuracy (OA) of 84.65% and a kappa accuracy (KA) of 82.60%. Based on overall accuracy and high kappa accuracy, the maps produced facilitate local governments and cooperatives in planning specific interventions for coffee-producing areas, supporting policies related to sustainable agriculture, climate-smart agroforestry expansion, and supply chain traceability.

Introduction

Bandung Regency is a region that contributes to Arabica coffee production, with 8,246.31 tons produced in 2022 [1]. According to the BPS [2] of Bandung Regency, the area of land planted with coffee is 13,853 hectares and will continue to increase in the future. Arabica coffee plantations in Bandung Regency are grown by farmers in forest areas permitted by the State-owned forestry company (Perum Perhutani) through *Pengelolaan Hutan Bersama Masyarakat* (PHBM). As market demand for coffee commodities increases, coffee plantations in Bandung Regency have begun to be grown on farmers' private land. Agricultural expansion is a primary driver of deforestation in tropical regions, including West Java. Studies indicate that 90 to 99% of deforestation in the tropics is associated with agriculture, although only 45 to 65% of deforested land becomes productive agricultural land [3]. Due to the limitation of forest area, while the need for agricultural land is continuously increasing, the state-owned forest corporation (Perum Perhutani) collaborated with the local community to utilize the area for agroforestry initiatives, which can help balance agricultural needs with forest conservation efforts. This collaborative approach may enhance sustainable land management while addressing the socio-economic needs of local populations.

Agroforestry systems, such as those involving cocoa and teak, are more productive and profitable than monoculture systems, suggesting that integrating trees into agricultural landscapes can support food security [4]. These collaborative forest management programs, such as those in West Java, have demonstrated potential in enhancing both forest ecological conditions and rural livelihoods by engaging local communities

Corresponding Author: I Nengah Surati Jaya ins-jaya@apps.ipb.ac.id Department of Forest Management, Faculty of Forestry, IPB University, Bogor, Indonesia.

in sustainable forest management practices [5]. The development of coffee commodities faces constraints as land use continues to change, raising new issues regarding how to allocate the available land for coffee cultivation among various existing land uses. Coffee agroforestry development in Bandung Regency has high potential, as it is one of the best producers of arabica coffee in Indonesia. The high potential of coffee production in Bandung Regency must be balanced with a comprehensive utilization plan to ensure it is optimally leveraged. However, there is an unavailability of data on the spatial distribution of coffee plantations, both those grown under agroforestry and monoculture systems.

Tabular data on the area of coffee plants published by central agencies differ from data found in regional agencies. Data on the potential of coffee commodities in Bandung Regency is still mixed with data on other commodities [6]. Therefore, the study aims to develop an algorithm that provides accurate spatial information (maps) for both coffee plantations in agroforestry and monoculture systems. Mapping of coffee plantation areas is necessary to achieve data harmony between the central and regional governments, enabling them to map and develop the existing potential effectively [7]. The use of Landsat 8 imagery, training datasets, predictors, and the DTML algorithm provides the most reliable classification method in detecting coffee agroforestry and monoculture in Bandung Regency. This study creating a novel classification approach for detecting coffee plantations, particularly agroforestry systems that are typically difficult to identify under dense canopy cover. The study produces the first reliable spatial dataset that separates coffee agroforestry from monoculture plantations in Bandung Regency. These data fill a critical gap, as previous government datasets aggregated coffee with other commodities and exhibited discrepancies between central and regional records.

The availability of spatial information, geophysical data, as well as coffee agroforestry and monoculture distribution maps, is reported. Therefore, the results serve as a reference in preparing spatial utilization plans and developing coffee agroforestry at the regional level in Bandung Regency. The resulting mapping methodology offers a scalable and replicable framework that supports data harmonization—allowing central and regional agencies to align their reported coffee plantation areas using verifiable spatial outputs rather than solely administrative statistics. In line with the development of information technology, satellite imagery-based land cover mapping is becoming increasingly straightforward to conduct because medium-resolution data are freely available and easily accessible, offering high-quality results. Previous study highlights the difficulties of mapping coffee agroforestry systems due to spectral similarity between coffee plants and forest canopies, the structural complexity of multilayered agroforestry arrangements, and varying topographic conditions [8,9]. The studies note that these factors limit the performance of optical imagery, which often struggles with spectral mixing and terrain-induced distortion.

To overcome these constraints, a semi-automatic method used to detect agroforestry and monoculture coffee plants is satellite imagery-based, combined with decision tree machine learning (DTML). This study emphasize the advantages of Landsat 8 imagery, which can penetrate cloud cover and capture structural information that enhances the accuracy of coffee coverage mapping. This can be applied in data analysis with the advantages of the decision tree method, including being easy to understand, practical, simple, and efficient, as well as applied in various data processing platforms [10]. Based on the problems mentioned earlier, this study was conducted to develop a DTML algorithm with a primary focus on assessing coffee plantations with agroforestry and monoculture systems in Bandung Regency.

The integration of various geophysical factors includes spatial data of land cover, elevation, slope, road proximity, and river proximity in Bandung Regency, combined with spectral variables in the form of vegetation index. Socioeconomic factors such as road distance, distance from settlements, and proximity to rivers significantly influence the site selection of coffee plantations in developing countries. These factors affect the accessibility, economic viability, and sustainability of coffee farming, which are crucial for optimizing production and ensuring the livelihoods of farmers [11,12]. The development plan for coffee agroforestry in Bandung Regency requires a comprehensive analysis of commodity-specific classification methods. This certainly raises the urgency to develop a reliable and accurate decision tree algorithm for land use classification at the commodity level, specifically coffee agroforestry.

Materials and Methods

The assessment of coffee agroforestry systems using DTML involves a structured workflow that integrates remote sensing data and socio-geobiophysical attributes. The process begins with (1) Landsat-8 image acquisition having characteristics as depicted in Table 1, followed by the (2) collection of supporting datasets, including socio-economic, geophysical, and biophysical variables. Subsequently, (3) creating synthetic images through the derivation of spectral indices (e.g., NDVI, NDBI, NDWI, and NRGI), and (4) creating socio-geobiophysical variables to enhance the predictive capacity of the model. The further steps are (5) a class scheme development focused in distinguishing between agroforestry-based coffee systems and monocultures, (6) delineation of training areas and calculation of zonal statistics, which provide representative samples for model development and validation, (7) attribute selection using four criteria: Brute Force (BF), Information Gain (IG), Gini Index (GI), and Gain Ratio (GR), (8) DTML model construction based on the selected attributes. The resulting model undergoes accuracy testing (9) using established validation metrics, and the best-performing model is subsequently selected (10). Finally, the validated model is applied to produce a spatial distribution map that differentiates between coffee agroforestry systems and monoculture plantations, offering valuable insights for sustainable land-use planning and agroforestry management. The study workflow is visually depicted in Figure 1.

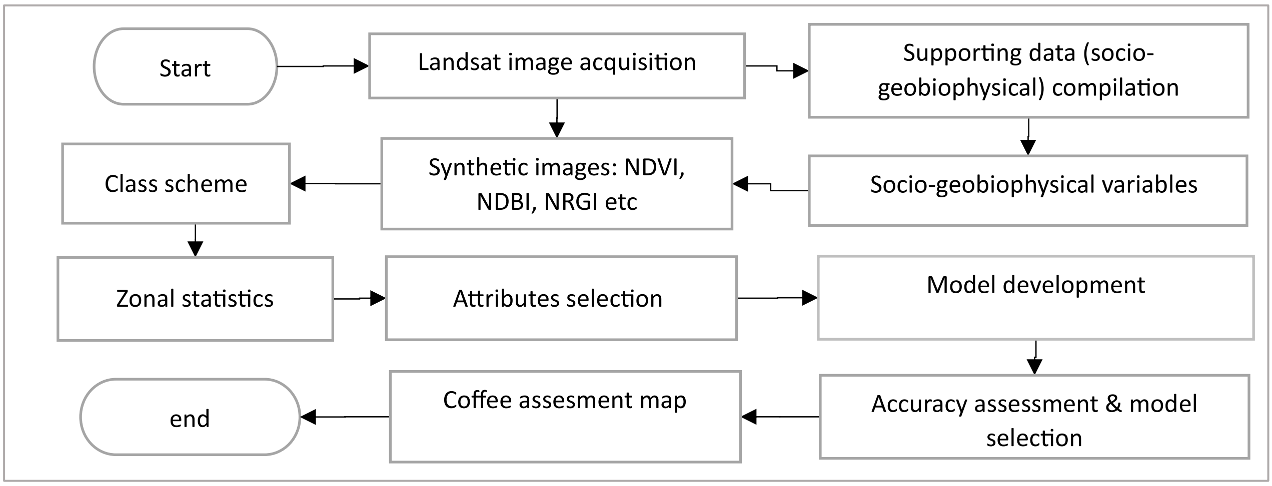


Figure 1. Sequential data processing steps used to derive coffee distribution information from Landsat 8, starting with satellite image acquisition and the compilation of supporting socio-geobiological datasets. Synthetic indices such as NDVI, NDBI, and NRGI are generated, combined with socio-geobiological variables, and integrated into a classification scheme. Zonal statistics and attribute selection are then applied to support model development, followed by accuracy assessment and model selection. This workflow results in the production of coffee assessment maps, illustrating a systematic and replicable approach to mapping coffee distribution in the study area.

Study Area

This study was conducted through the stages of measurement and data survey in Bandung Regency, located between 107°14'–107°56' East Longitude and 06° 49'–07°18' South Latitude. Data collection in the field used the purposive sampling method for Cimaung District, Ciwidey District, Ibun District, Kertasari District, Pangalengan District, Paseh District, and Rancabali District (Figure 2). A total of 60 sample plots (2,056 pixels) were spread across the district, representing coffee agroforestry and coffee monoculture land cover.

Data Collection and Analysis

Landsat 8 imagery provides a comprehensive set of spectral bands that are widely used for environmental monitoring, land use/land cover mapping, and resource assessment. Each band is designed to capture information from specific portions of the electromagnetic spectrum, with spatial resolutions of either 30 meters for multispectral bands or 15 meters for the panchromatic band. The coastal/aerosol band (Band 1) is useful for coastal and aerosol studies. The visible bands consisting of blue (Band 2), green (Band 3), and red (Band 4) are essential for natural color composites, vegetation studies, and water body analysis. The near-infrared (NIR, Band 5) is particularly valuable for vegetation health and biomass monitoring due to its sensitivity to plant reflectance. Shortwave infrared bands (SWIR-1, Band 6, and SWIR-2, Band 7) enhance the detection of soil and vegetation moisture as well as geological features. The panchromatic band (Band 8) offers finer spatial detail at 15 meters, enabling sharper image interpretation and pan-sharpening techniques. Finally, the cirrus band (Band 9) is designed to detect high-altitude clouds that may affect image analysis (Table 1).

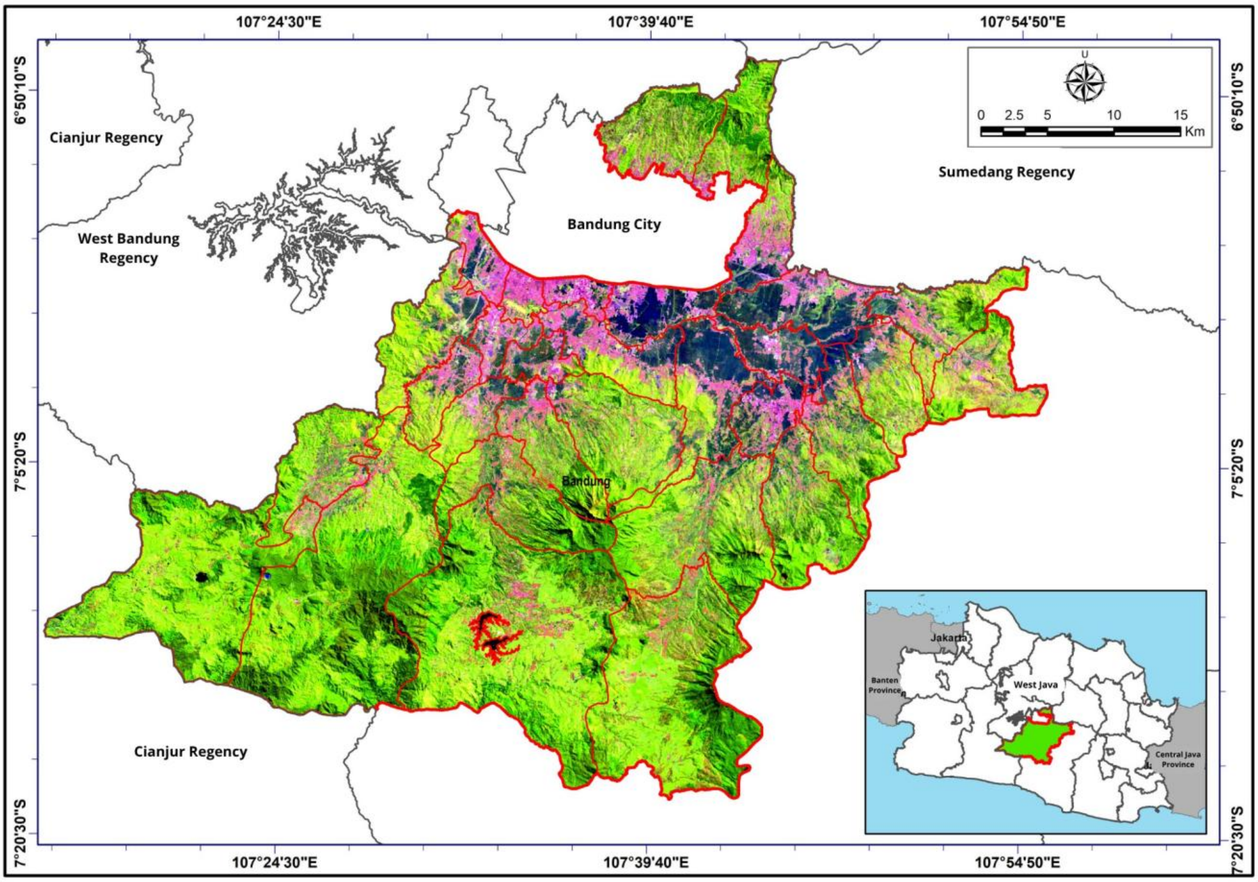


Figure 2. Spatial distribution of land cover and administrative boundaries in Bandung Regency, West Java, Indonesia. The map depicts Bandung Regency (outlined in red), its surrounding by other administrative areas (Bandung City, West Bandung Regency, Cianjur Regency, and Sumedang Regency), and major landscape features including elevation gradients and built-up areas. Color variations indicate differences in land cover, with dense urban zones concentrated in the northern part of the regency. The inset map highlights the location of Bandung Regency within West Java Province. Overall, the figure illustrates the strong contrast between highly urbanized northern areas and the predominantly rural, vegetated landscapes in the central and southern regions.

Table 1. Summarizes of Landsat 8 imagery, including their band designations, spatial resolutions, and wavelength ranges. The listed bands cover key portions of the electromagnetic spectrum from coastal/aerosol to visible, near-infrared (NIR), shortwave infrared (SWIR), panchromatic, and cirrus, each serving distinct analytical purposes such as vegetation monitoring, land-use mapping, moisture detection, and atmospheric correction. Table highlights the multispectral capability of Landsat 8 for environmental and land-use analysis, including vegetation assessment and coffee distribution mapping.

Imagery type	Channel sensor	Spatial resolution (m)	Wavelength (µm)
Landsat 8 imagery	Band 1: Coastal/Aerosol	30	0.435 – 0.451
	Band 2: Blue	30	0.452 – 0.512
	Band 3: Green	30	0.533 – 0.590
	Band 4: Red	30	0.636 – 0.673
	Band 5: NIR	30	0.851 – 0.879
	Band 6: SWIR-1	30	1.566 – 1.651
	Band 7: SWIR-2	30	2.107 – 2.294
	Band 8: PAN	15	0.503 – 0.676
	Band 9: Cirrus	30	1.363 – 1.384

Note: NIR = near infrared, PAN = panchromatic, SWIR = short wave infrared.

Several measuring instruments were used, such as a Global Positioning System (GPS), a camera, a phi-band, a clinometer, a compass, a tally sheet, a measuring tape, a rope, and stationery. In addition, data processing and analysis were carried out using qantum GIS software, Codeblocks 20.03, and rapid miner studio. Primary data was Landsat 8 imagery, path 121 – row 065, with a resolution of 30 × 30 m recorded on 26 July 2022 (Table 1) and field data. The Landsat 8 imagery was obtained from the United States Geological Survey (USGS) website.

Secondary data were in the form of geophysical factors, such as spatial data on elevation, slope, road proximity, and river proximity in Bandung Regency. These data were downloaded from Data DEMNAS and Data RBI of Geospatial for the Country of Indonesia (https://tanahair.indonesia.go.id/portal-web/unduh accessed on 26 December 2022). Data pre-processing was carried out to examine satellite imagery data before conducting further analysis. The procedure started with cloud correction on Landsat 8 imagery, class scheme creation, training area creation, and synthetic creation (Table 2).

Table 2. Structured pre-processing in ensuring the reliability and analytical value of satellite-derived products, particularly for applications coffee distribution assessment, including cloud correction, class scheme creation, training area development, and the generation of synthetic indices. For each step, the table outlines the required input data, the methods applied such as cloud masking using QA Pixels, visual interpretation of spectral composites, and mathematical transformations of spectral bands, and the resulting outputs. The listed outputs, including class schemes, training area datasets, and synthetic indices e.g., NDVI, EVI, NBR, and related vegetation and moisture metrics, demonstrate how multispectral information from Landsat 8 is systematically processed to support accurate land-cover mapping.

Stages	Pre-processing
Cloud correction of Landsat 8 imagery	**Input:** Landsat 8 imagery recorded in 2022. **Method:** Cloud masking algorithm Quality Assessment Pixel (QA Pixel). **Output:** Cloud-free Landsat 8 imagery.
Creating class schemes	**Input:** Cloud-free Landsat 8 imagery recorded in 2022. **Method:** Proximity of reflectance values, spectral and spatial observations of cloud-free Landsat 8 imagery. **Output:** Scheme of classes, including natural forest, plantation forest, dryland agriculture and shrubs, plantations, rice fields, water bodies, open land, settlements, coffee agroforestry, and coffee monoculture.
Creating training areas (TA)	**Input:** Cloud-free Landsat 8 imagery recorded in 2022. **Method:** Interpretation using visual methods using a combination of Red-Green-Blue and NIR-Red-Green channels of Landsat 8 imagery. **Output:** 2056 pixels of TA, including 190 for natural forest, 177 for plantation forest, 348 for dryland agriculture and shrubs, 200 for plantations, 270 for paddy fields, 101 for water bodies, 130 for open land, 328 for settlements, 177 for coffee agroforestry, 135 for coffee monoculture. 70% of the TA was allocated for model development, while the remaining 30% was used for model validation (accuracy assessment)
Creating synthetic imagery	**Input:** Cloud-free Landsat 8 imagery recorded in 2022. **Method:** Mathematical operations on pixels in Landsat 8 imagery channels. **Output:** Atmospheric reflection vegetation index, enhanced vegetation index, green atmospherically resistant index, modified normalized differences wetness index, green-based, normalized red-green vegetation index, and visible difference vegetation index.

The study performed several preprocessing steps to ensure data quality and accuracy. Preprocessing typically involves correcting atmospheric and cloud-related disturbances, since cloud contamination is a common limitation in optical satellite data. In this study, cloud correction was conducted using the Quality Assessment (QA Pixel) algorithm to generate a cloud-free image mosaic of the study area, ensuring more reliable spectral information for subsequent analysis. A series of processing stages was implemented to prepare the imagery for classification. These included the development of land cover class schemes based on spectral and spatial characteristics, the delineation of training areas through visual interpretation of composite images (RGB and NIR combinations), and the generation of synthetic imagery using mathematical transformations of spectral bands to derive vegetation and wetness indices.

In this study, we applied a decision tree (DT) machine learning algorithm, a widely recognized non-parametric classification method in remote sensing [13,14], to evaluate whether its application can enhance classification accuracy despite the moderate spatial resolution of Landsat imagery. The approach emphasizes the integration of both spectral indices and geophysical variables, with the expectation that combining these data sources may improve the separability of land cover classes and support more robust mapping results. The synthetic images or image indices developed in this study were images that were derived from mathematical operations using pixels originating from several channels in the imagery [15]. Previous study showed that the channels influential in vegetation detection were visible waves (red, green, and blue) as well as near-infrared and mid-infrared [16,17]. This study develops a coffee agroforestry detection algorithm using another method through the Atmospheric Reflection Vegetation Index (ARVI), Enhanced Vegetation Index (EVI), Green Atmospherically Resistant Index (GARI), Modified Normalized Differences Wetness Index, green-based (MNDWIg), Normalized Red-Green Vegetation Index (NRGI), and Visible Difference Vegetation Index (VDVI).

NDVI is widely used for monitoring vegetation dynamics due to its ability to indicate photosynthetic activity and vegetation greenness, as well as its sensitivity to changes in vegetation cover and environmental degradation assessment [18]. Meanwhile, ARVI is designed to be less sensitive to atmospheric effects compared to NDVI. It incorporates the blue channel to correct atmospheric influences on the red channel, enhancing its reliability in varying atmospheric conditions. This makes ARVI particularly useful in regions with frequent atmospheric disturbances, as it provides more accurate vegetation monitoring. GARI is similar to ARVI but focuses on the green spectrum, offering improved resistance to atmospheric effects. It is beneficial in environments where atmospheric conditions can significantly impact remote sensing data [19].

The MNDWI is used to monitor water bodies and moisture content in vegetation. It is effective in detecting changes in water surface areas and vegetation moisture, making it valuable for assessing environmental degradation and managing water resources. MNDWI has been applied in studies to track changes in water bodies and wetlands, providing insights into hydrological dynamics and the impacts of land use [20]. The NRGI is used to assess vegetation health by analyzing the red and green spectral bands. It plays a crucial role in distinguishing between different types of vegetation and assessing their health status [21]. The VDVI is designed to enhance the sensitivity of vegetation monitoring, particularly in areas with high vegetation density. It provides a more dynamic range of vegetation detection, improving the accuracy of monitoring efforts [22]. Mathematical operations for creating synthetic imagery with Landsat 8 imagery include the following equations:

$$ARVI = \frac{(NIR - (2 \times Red - Blue))}{(NIR + (2 \times Red - Blue))} \tag{1}$$

$$EVI = \frac{2.5\ (NIR - Red)}{(NIR - 6Red + 7.5Blue + 1)} \tag{2}$$

$$GARI = \frac{NIR - (Green - 1.7\ (Blue - Red)}{NIR + (Green - 1.7\ (Blue - Red)} \tag{3}$$

$$MNDWIg = \frac{Green - SWIR}{Green + SWIR} \tag{4}$$

$$NRGI = \frac{Green - Red}{Green + Red} \tag{5}$$

$$VDVI = \frac{((2 \times Green) - Red - Blue)}{((2 \times Green) + Red + Blue)} \tag{6}$$

This study uses five vegetation and wetness indices to detect cover and water bodies with unique characteristics (Figure 3). Specific land covers produce reflections with different wavelengths known as hues. Meanwhile, vegetation cover hues in synthetic imagery, such as ARVI, EVI, GARI, NRGI, and VDVI, produce bright hues. Land covers in the form of water bodies, open land, and settlements produce darker hues. Synthetic imagery MNDWIg for wetness index shows brighter hues for water bodies, while other land covers possess darker hues.

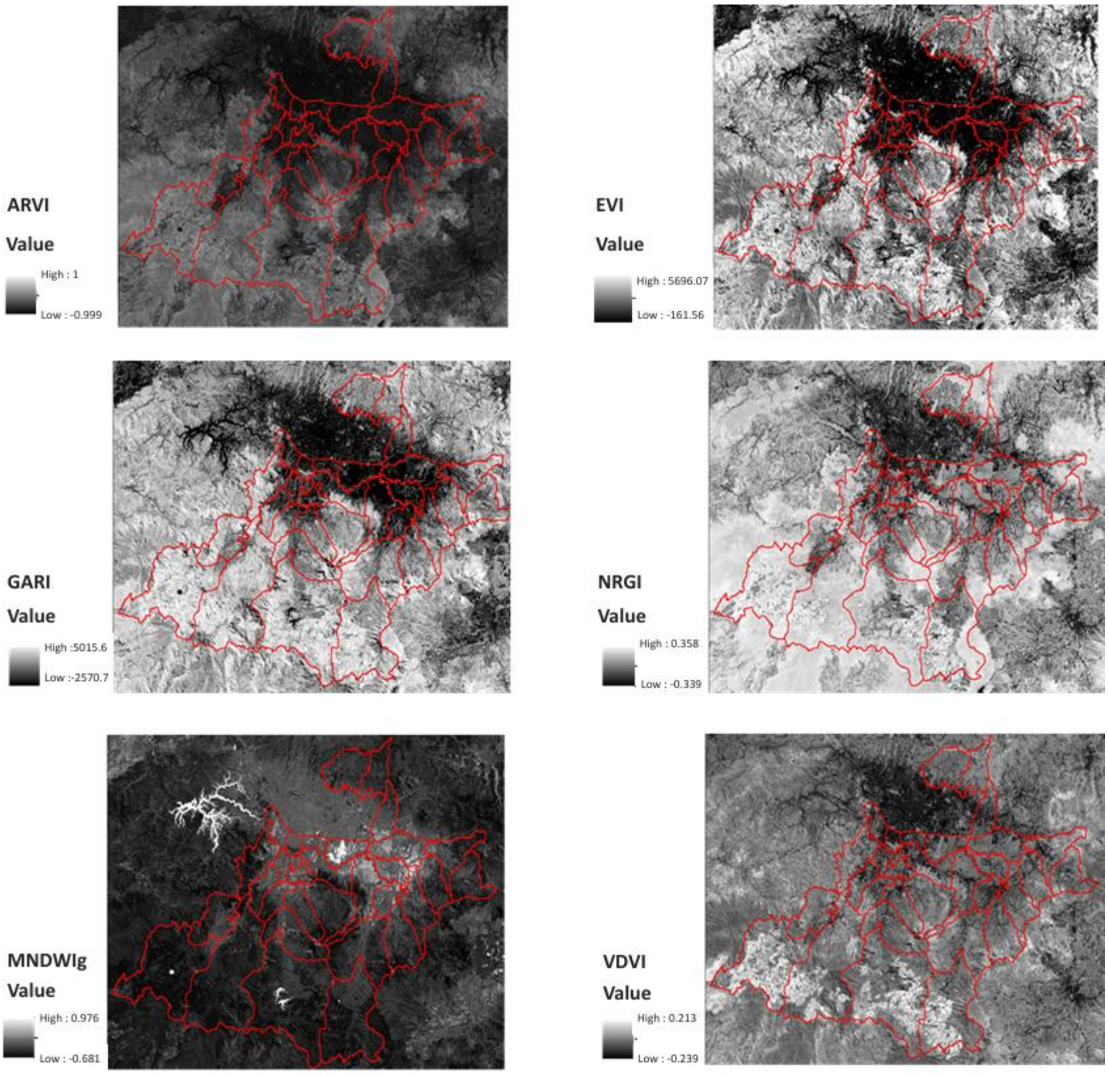


Figure 3. Synthetic vegetation-index maps derived from Landsat 8 imagery for the year 2022, displaying ARVI, EVI, GARI, NRGI, MNDWIg, and VDVI across the study area with administrative boundaries overlaid in red. These panels illustrate spatial variations in vegetation density, greenness, moisture, and spectral responses, enabling comparison of how different indices represent ecological conditions. Figure highlights consistent patterns of higher vegetation activity in darker-toned regions, demonstrating the complementary value of multiple indices for characterizing landscape-level vegetation dynamics.

Decision Tree Machine Learning Analysis

A decision tree is a classification method used for the induction of machine learning algorithms. The method uses a tree structure with the root, branch (sub-trees), and leaf nodes [23]. The decision tree method has been effectively utilized for image classification in land-use and land-cover research [24]. A decision tree is compiled using Rapidminer software with the C4.5 algorithm, starting with calculating the entropy value. Based on the description, the construction is performed by dividing data into several small groups with the same value [25]. The division should decide which variables are the best and most predictive. The creation of a decision tree hierarchy uses criteria from the variables known as the parameters. A decision tree is created by dividing the training data set into several different nodes [26].

The measurement of the algorithm starts by calculating the entropy value. In this context, entropy is defined as the sum of the probabilities of each label multiplied by the log probability of the label. Entropy value measures the impurity or level of randomness in a data set [27]. The equation for calculating entropy is as follows:

$$Entropy\ (S) = -\sum_{i=1}^{n} - P_{\mathrm{i}}\ log_2 P_{\mathrm{i}} \qquad (7)$$

Description:

Pi = proportion of data

S = total data set

n = number of data in S

Split info value is calculated to state the entropy or impurity of a variable. This value is obtained from the sum of the proportions of sub-attributes using the following equation:

$$Split\ Info\ (S,A) = -\sum_{i=1}^{k} \frac{S_i}{S}\ log^{2\frac{S_i}{S}} \quad (8)$$

Description:

S = total data set

Si = number of data in sub-variable i

Entropy shows the calculation for obtaining information gain, which is the difference between class entropy, conditional class entropy, and the selected variable. Information gain is used to measure the change in entropy before and after class separation [28]. The variable with the highest value is selected for separation, as reported in the following equation:

$$Information\ Gain = Entropy\ (S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} Entropy\ (S_i) \quad (9)$$

Description:

|Si| = number of cases of partition i

|S| = number of cases in S

Si = number of data in sub-variable i

The ranking of variables with gain ratio is used to evaluate the level of importance of all n variables. The measure is the result of dividing the information gained by the split info.

$$Gain\ Ratio = \frac{Gain\ (S,A)}{Split\ info(S,A)} \quad (10)$$

Description:

S = case set

A= variable

The Gini index is used to determine the purity of a particular class after separation based on a particular variable [29]. The best separation increases the purity of the set produced. The Gini index equation is defined below when K is a dataset with j being a different class label.

$$Gini\ Index = 1 - \sum_{i=1}^{n} - P_{i^2} \quad (11)$$

Description:

Pi = data proportion of class-i

The application of big data analytics to spatially referenced data in environmental monitoring presents both significant challenges and opportunities. The challenges primarily stem from the complexity and heterogeneity of environmental data, while the opportunities lie in the potential for enhanced predictive capabilities and decision-making processes. The integration of advanced technologies such as cloud computing and machine learning can further enhance the utility of big data analytics in this field. This big data often contains a certain amount of noise and outliers. A decision tree is a method that is sensitive to noise and outliers. Therefore, one of the disadvantages of the decision tree method is overfitting. Overfitting occurs because the resulting decision tree model is too complex, so that the model will adjust to the training data. To overcome this, additional methods should be applied to the decision tree to reduce overfitting and improve accuracy.

Methods that can be used to overcome overfitting include pruning. Pruning is emphasized by removing unnecessary nodes in the decision tree [30]. The pruning process is done in two ways: the first is to use pre-pruning, which is to stop the splits so that the decision tree does not grow larger, which means it will stop the tree at a specific growth rate. The second, referred to as final pruning, removes split nodes created after the tree is fully formed.Pruning is an effort to reduce the number of errors generated in the results of land cover classification with a decision tree. Previous studies have shown that pruning can effectively improve accuracy by up to 11% [31]. Other studies show that overfitting can be reduced by pruning and result in a practical accuracy value of about 92% [32]. Another method to reduce overfitting is to use brute force, which serves to prune the amount of data that is not influential or not relevant. The brute force method has the principle of selecting the best set of variables by testing all possible combinations of the selected variables. The brute force combination is C_k^j, j being the total number of variables and k being the possible number of selected variables. Irrelevant or distracting variables often confuse the decision tree machine learning process, resulting in models with low accuracy. In addition, selecting variables can result in a simpler model that uses less data and has the capability for land cover classification [33].

Optimization of Model Selection and Parameters

Information gain, Gini index, gain ratio, and brute force are some of the variable selection and ranking methods used. In this context, information gain is based on entropy to measure the usefulness of variables [34]. The Gini index determines the purity of a class after separation based on a particular variable. Ranking with a gain ratio is used to evaluate the level of all variables. Meanwhile, Brute force is a method used to select the most relevant variables, where a value of 1 shows usage in the classification. The results of Brute force calculation are in the form of a weight value for each variable used. The straightforward method is used to solve problems requiring input and consideration [35].

The analysis is continued with the optimization of model parameters to achieve the best classification accuracy. The criteria used are decision tree, maximal depth, minimal leaf size, pre-pruning, and pruning. The development of decision tree models often experiences obstacles in the form of data overfitting. The method used to overcome data overfitting is the use of pre-pruning and pruning [36].

Accuracy Test

The accuracy test analyzes the level of success in classifying objects using the decision tree method. Commonly used types are overall accuracy (OA) and kappa accuracy (KA). The accuracy test uses the principle of the confusion matrix as a comparison that contains reference and classification result data.

Results

Selection of the Best Variables and Models

Visual-based Land cover, spectral, and geophysical variables were combined to develop the best model for detecting the spatial distribution of coffee plants. The selection of variables was carried out by calculating the weight of each variable, as reported in Tables 3, 4, 5, and 6. The weight value shows a higher relevance level of the variable. The magnitude affects the amount of information in distinguishing objects within a particular class. The weight of each variable in Bandung Regency for the geophysical combination can be seen in Table 3.

Table 3. Relative weights of four geophysical variables such as elevation, slope, road proximity, and river proximity, calculated using information gain, Gini Index, gain ratio, and a Brute Force Method. These weighting schemes are applied to assess the influence of each variable within a geospatial analysis framework, enabling comparison across different feature-selection metrics. The results consistently identify elevation and slope as the most influential predictors, while road and river proximity exhibit minimal contribution across all weighting methods showed by zero value in information gain, Gini Index, and gain ratio.

Weight							
Variable	Information gain	Variable	Gini index	Variable	Gain ratio	Variable	Brute force
Elevation	1.000	Elevation	1.000	Elevation	1.000	Elevation	1
Slope	0.495	Slope	0.656	Slope	0.646	Slope	1
Road proximity	0.078	Road proximity	0.119	Road proximity	0.020	Road proximity	1
River proximity	0	River proximity	0	River proximity	0	River proximity	1

Geophysical variables provide information about the condition of a land. The uniqueness and characteristics of the landscape can be seen from the inherent geophysical variables [37]. Table 3 shows that elevation is the most influential variable in the model, with information gain, Gini index, gain ratio, and brute force values of 1. Therefore, the root node in the geophysical combination is the elevation variable used as the starting point for dividing the dataset into homogeneous and smaller subsets. Elevation variables can separate land cover classes in Bandung Regency with lower entropy values, so that the separation will result in a more homogeneous class.

Elevation is the most influential variable because the conditions in Badung Regency have significant differences due to the presence of mountains, highlands, and lowlands. Elevation provides the most accurate value among other geophysical variables to detect agroforestry coffee land cover based on the habit of agroforestry coffee planting by the farmers in Bandung Regency, which is easily found at high elevations, usually more than 1,500 meters above sea level (masl). In addition, based on land suitability data for arabica coffee cultivation in Bandung Regency, the elevation range is 1,500 to 2,000 masl. Slope is the second ranking in Table 3, which shows farmers' habits for growing coffee in Bandung Regency. Coffee monoculture will be planted on gentle to steep slopes, ranging from 8% up to 25%. For accessibility reasons, in transporting coffee bean production and management, both coffee agroforestry and monoculture were planted close to the road [38].

River proximity does not affect the spatial distribution of coffee plantations, as the water requirement of coffee does not require regular irrigation and watering. In Bandung Regency, the microclimate created by the surrounding vegetation, such as pine trees, helps maintain stable humidity levels, which is beneficial for coffee cultivation [39]. That is why the coffee plantation does not need direct water irrigation from the river. Rainfall is another critical factor, as it provides the necessary water supply for coffee plants. The ideal annual rainfall for Arabica coffee ranges from 1,800 to 2,000 mm, which aligns with the rainfall patterns observed in Bandung Regency [40]. Consistent rainfall ensures that coffee plants receive adequate water throughout the year, reducing the need for artificial irrigation and supporting sustainable coffee farming practices [41]. Elevation significantly impacts coffee plantation suitability due to its influence on temperature and microclimate. Arabica coffee thrives at elevations between 1,000 and 1,500 masl, where temperatures are cooler and more stable [42,43]. In Bandung Regency, the elevation provides a favorable environment for coffee growth by creating a microclimate that reduces the risk of pests and diseases, which are more prevalent at lower altitudes [44].

In contrast, the proximity of roads still influences coffee cultivation. The coffee cultivation in Bandung Regency, both in agroforestry and monoculture systems, is mostly close to the road network, meeting the needs of farmers for crop management, harvesting, and distribution. As described in Table 4, the spectral variables used produce varying effects, and none is the most dominant. ARVI shows the highest information gain value, while the best NRGI values for both the Gini index and the gain ratio are 1. The ARVI attribute is the root node in the model with a value of 1 and 0.850 for information gain and gain ratio, respectively. The attribute is a good vegetation index in Landsat 8 imagery due to resistance to atmospheric effects [45]. These spectral variables in Table 4 provide a better visual representation of the coffee plants grown in Bandung Regency.

Table 4. The relative weights of ARVI, EVI, NRGI, GARI, MNDWIg, and VDVI indicated that ARVI, EVI, and NRGI consistently have the highest influence, with values close to 1, while MNDWIg and especially VDVI contribute minimally to the spectral combination analysis with values close to 0.

Weight							
Variable	Information gain	Variable	Gini index	Variable	Gain ratio	Variable	Brute force
ARVI	1.000	NRGI	1.000	NRGI	1.000	ARVI	1
EVI	0.645	VDVI	0.773	ARVI	0.850	EVI	1
NRGI	0.496	ARVI	0.670	EVI	0.285	NRGI	1
GARI	0.254	EVI	0.484	GARI	0.238	GARI	1
MNDWIg	0.084	GARI	0.224	MNDWIg	0.182	MNDWIg	1
VDVI	0	MNDWIg	0	VDVI	0	VDVI	1

The combination of red-blue (RB) channels in ARVI attribute minimizes the atmospheric scattering effect caused by aerosols in the red channel [46]. In addition to the ARVI variable, the EVI and GARI variables are also well used in Landsat 8 imagery because of their ability to correct for atmospheric effects. Another variable, MNDWIg, is a wetness index used to detect water bodies on the earth's surface [47]. VDVI is a vegetation index similar to NDVI, a development that has a similar range of values. Other spectra, such as EVI, GARI, MNDWIg, and VDVI, remain relevant for use in the model. This is because brute force is 1, even though VDVI is 0 in information gain and gain ratio, as well as 0 for MNDWIg in the Gini index (Table 4). Therefore, the six variables are still used in spectral combination to build a decision tree algorithm.

Observations by combining spectral and geophysical data were conducted to increase the retrieval of more information in the model built. Spectral variables have a greater influence on the decision tree model compared to geophysical factors, as shown in Table 5, which indicates that spectral variables ranked among the top 6. Spectral variables provide better information on land conditions compared to geophysical variables. ARVI is the root node in the decision tree with information gain and gain ratio of 1 and 0.950, respectively (Table 5). Spectral-geophysical combination increases the values of each variable in IG, GI, and GR. The combination of ancillary data and spectral increases the value of information and influence because the variables cannot necessarily distinguish land cover well [48]. The variables have a Brute force value of 1, which is relevant and influential in the decision tree model (Table 5).

Table 5. Relative weights of spectral and geophysical variables derived from four feature-selection criteria to evaluate their importance in environmental or remote-sensing–based analyses. The table shows that spectral indices consistently receive the highest weights across all methods, indicating their dominant contribution to model performance compared with geophysical variables such as elevation, slope, and proximity measures.

Weight							
Variable	Information gain	Variable	Gini index	Variable	Gain ratio	Variable	Brute force
ARVI	1.000	NRGI	1.000	NRGI	1.000	ARVI	1
EVI	0.827	VDVI	0.881	ARVI	0.950	EVI	1
NRGI	0.755	ARVI	0.828	EVI	0.761	NRGI	1
GARI	0.637	EVI	0.731	GARI	0.746	GARI	1
MNDWIg	0.555	GARI	0.596	MNDWIg	0.727	MNDWIg	1
VDVI	0.514	MNDWIg	0.479	VDVI	0.666	VDVI	1
Elevation	0.423	Elevation	0.327	Elevation	0.245	Elevation	1
Slope	0.209	Slope	0.215	Slope	0.158	Slope	1
Road proximity	0.033	Road proximity	0.039	Road proximity	0.005	Road proximity	1
River proximity	0	River proximity	0	River proximity	0	River proximity	1

Another combination tested includes spectral, geophysical, and land cover data variables (Table 6). The root node in the decision tree model is land cover (PL Vis) with an Information Gain of 1, a Gini Index of 1, and a Gain Ratio of 0.419. Land cover has a high influence and relevance in the decision tree algorithm model. This is because the variable provides initial information on land cover classes. The combination of spectral-geophysical with PL Vis provides a model computation similar to supervised classification. The variables PL Vis, ARVI, EVI, GARI, MNDWIg, NRGI, VDVI, elevation, slope, road proximity, and river proximity affect the model characterized by a brute force value of 1. The combinations of geophysical, spectral, spectral-geophysical, and spectral-geophysical-land cover are optimized for each parameter to determine the best model in the decision tree.

Table 6. The relative importance weights of spectral, geophysical, and land-cover variables as determined by four feature-selection methods: information gain, Gini index, gain ratio, and brute force. The evaluated variables include vegetation indices, moisture indices, topographic attributes, proximity measures, and a land-cover parameter (PL Vis), representing a comprehensive set of environmental predictors. The results show that PL Vis consistently receives the highest weight across all methods, indicating that land-cover information plays a dominant role in explaining variability compared with other spectral and geophysical variables.

Weight							
Variable	Information gain	Variable	Gini index	Variable	Gain ratio	Variable	Brute force
PL Vis	1.000	PL Vis	1.000	NRGI	1.000	PL Vis	1
ARVI	0.419	NRGI	0.251	ARVI	0.950	ARVI	1
EVI	0.347	VDVI	0.222	EVI	0.761	EVI	1
NRGI	0.316	ARVI	0.208	GARI	0.746	GARI	1
GARI	0.267	EVI	0.184	MNDWIg	0.727	MNDWIg	1
MNDWIg	0.232	GARI	0.150	VDVI	0.666	NRGI	1

Weight							
Variable	Information gain	Variable	Gini index	Variable	Gain ratio	Variable	Brute force
VDVI	0.215	MNDWIg	0.120	PL Vis	0.419	VDVI	1
Elevation	0.177	Elevation	0.082	Elevation	0.245	Elevation	1
Slope	0.087	Slope	0.054	Slope	0.158	Slope	1
Road proximity	0.014	Road proximity	0.010	Road proximity	0.005	Road proximity	1
River proximity	0	River proximity	0	River proximity	0	River proximity	1

Land cover variables provide an initial overview of land use in Bandung Regency based on the mapping that has been created, but do not specifically show the distribution of coffee agroforestry and coffee monoculture. Spectral variables in the form of vegetation index can detect land cover in the form of vegetation and non-vegetation based on the greenness degree value. Bio-socio-geophysical variables provide information on traditional practices and physical characteristics of coffee planting patterns carried out by farmers, including suitable altitude (elevation), the slope selected for coffee planting, road access to the planting site (proximity to roads), and access to irrigation (proximity to rivers). Based on the results given in Table 6, the integration or combination of spectral variables with bio-socio-geophysical variables as additional information is proven to increase the percentage of success in land cover classification.

Best Model Parameter Optimization

Parameter optimization uses the iteration method to obtain a combination of 234,256. The optimal value is determined based on the overall accuracy value [49]. The parameter is selected based on the 10 best rankings from thousands of combinations. The information gain parameter achieves the highest overall accuracy value of 84.65% with an 8-fold selection in Table 7. Another parameter selected in the optimization is the Gini index, with an overall accuracy of 84.23% being the highest. The best information gain parameter requires a pruning process with a minimum leaf node of 11, a minimum of 31 samples, and a maximum tree depth of 80. Pruning in the optimal model is carried out to cut or remove several unnecessary branches with weak influences [50]. The process is carried out to develop the reliability and accuracy of the decision tree, which has been proven to increase overall accuracy [51].

Table 7. The performance of decision tree models optimized under different parameter configurations, including splitting criteria (IG or GI), pre-pruning and pruning options, minimal leaf and split sizes, and maximal tree depth. The results provide a comparative context for understanding how variations in pruning strategies and structural parameters influence classification accuracy. The table shows that several parameter combinations yield similar high accuracies around 84%, indicating that the model is relatively robust to these adjustments, with IG–based trees showing slightly better optimal performance.

Criteria	Pre-pruning	Pruning	Minimal leaf size	Minimal size for split	Maximal depth	Overall accuracy (%)
IG	F	T	11	31	80	84.65
IG	F	F	60	51	9	84.23
GI	F	F	31	70	39	84.23
IG	F	F	31	1	29	84.14
GI	F	T	41	1	50	84.14
IG	F	T	21	80	100	83.97
IG	F	F	11	90	100	83.97
IG	F	F	1	31	9	83.89
IG	F	F	100	60	19	83.89
IG	F	T	1	11	100	83.80

Note: F = false, GI = Gini index, IG = information gain, T = true.

Accuracy Test Results

The models built using different combinations produce varying levels of accuracy. Geophysical combination uses elevation, slope, road proximity, and river proximity, while spectral combination only adopts the vegetation index. Observations are made by combining geophysical and spectral variables, as well as geophysical-spectral-land cover. To address the potential risk of overfitting in the classification process, the reference samples were systematically divided into two groups. A total of 70% of the samples were allocated for model development (training), ensuring that the decision tree algorithm could learn the relationships between spectral and geophysical variables. The remaining 30% of the samples were reserved exclusively for validation, thereby providing an independent dataset to evaluate model performance. This separation of training and validation data is a standard practice in machine learning-based studies, as it enables a more objective assessment of classification accuracy and ensures that the results are not biased by the data used for model construction.

In addition, the confusion matrix is arranged to obtain OA and KA values for each combination. The geophysical combination provides the lowest results with OA and KA of 52.12% and 45.20%, respectively. We noticed that there is a discrepancy between the overall accuracy (84.65%) and the lower producers' accuracies for the coffee-related classes (coffee agroforestry/CAF= 68.27%, coffee monoculture/CNF = 57.33%). We acknowledge that while the classification achieved high overall performance, the confusion matrices indicate challenges in distinguishing between these spectrally and structurally similar classes. CAF and CNF often share overlapping spectral signatures, particularly in medium-resolution imagery such as Landsat (30 m), where mixed pixels can obscure subtle differences in canopy structure and understory conditions. Nonetheless, the inclusion of these classes remains important for our study's objectives.

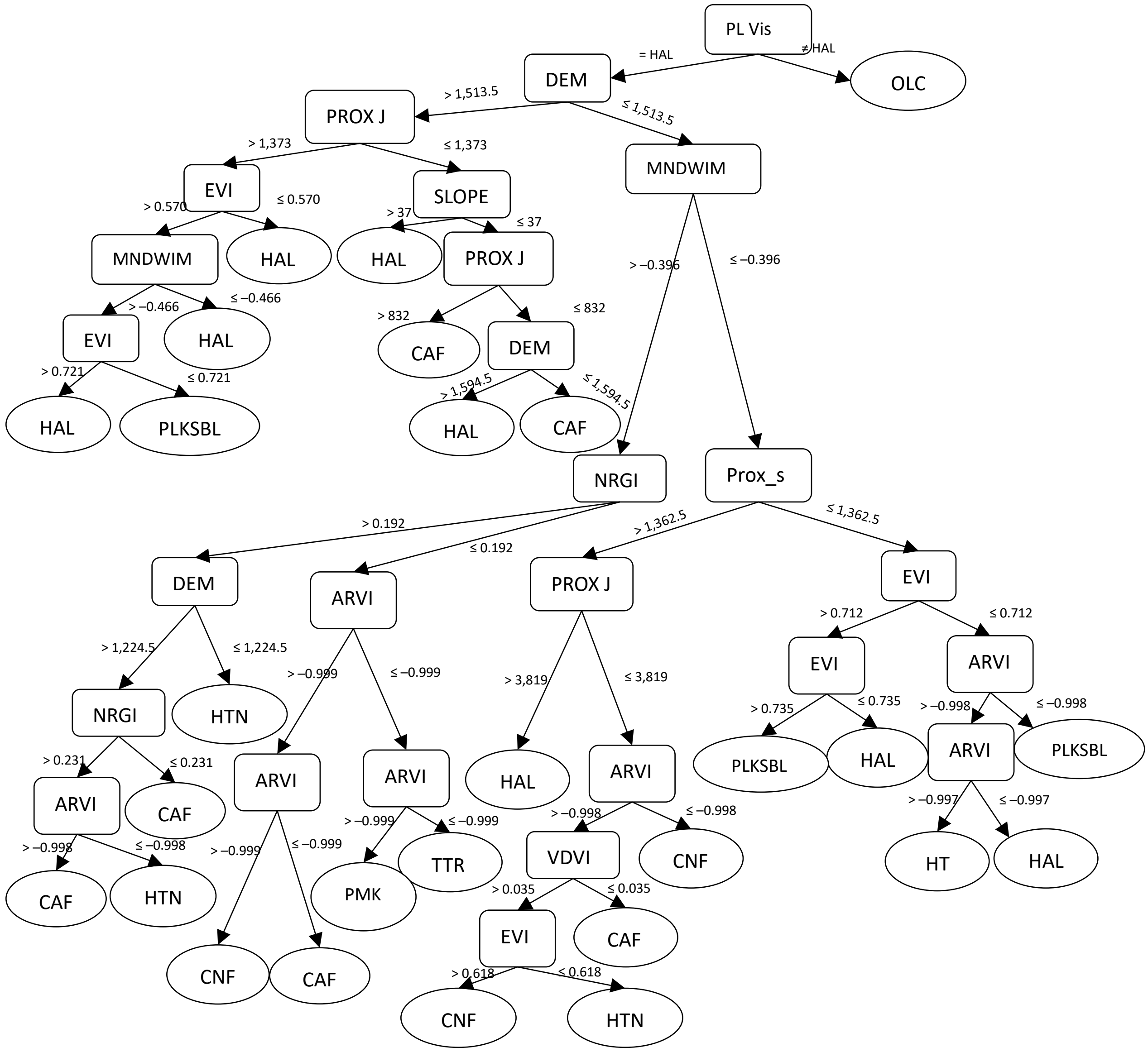
Our results highlight both the potential and the limitations of using Landsat imagery for discriminating coffee production systems. We interpret the relatively lower accuracies not as a shortcoming of the method, but as evidence of the inherent difficulty of this classification problem. In future work, we plan to address this limitation by integrating higher-resolution imagery and other additional ancillary variables (e.g., LiDAR, topography, or climate data), which may enhance class separability. Geophysical variables have low accuracy in providing additional information to describe the existence of Arabica coffee through habitual patterns [52]. The variables should be combined to obtain higher accuracy results. The combination of geophysical-spectral-land cover has the highest accuracy value compared to other combinations at 84.65% and 82.60% for OA and KA. Geophysical-spectral-land cover is the best decision tree algorithm selected from the four existing combinations (Table 8).

Table 8. The overall accuracy and Kappa accuracy values derived from models built using different combinations of geophysical, spectral, and land-cover variables. It provides a comparative context to assess how each variable group or combination influences classification performance. The results show that integrating geophysical, spectral, and land-cover variables yields the highest accuracies, with OA 84.65% and KA 82.60%, indicating that multi-variable models perform substantially better than those relying on single variable groups.

Combination of variables	Accuracy test results (%)		Selected variable
	Overall accuracy (OA)	Kappa accuracy (KA)	
Geophysical	52.12	45.20	Elevation
Spectral	77.49	74.50	ARVI
Geophysical-spectral	80.05	77.40	ARVI
Geophysical-spectral-land cover	84.65	82.60	PL Vis

The results of the accuracy test found that the combination of geophysical-spectral variables-land cover, with an OA of 84.65% was able to provide the best decision tree algorithm to be used in land cover classification, especially to detect agroforestry and monoculture coffee in Bandung Regency. The decision tree algorithm with this combination can be used to create a reliable map of agroforestry and monoculture coffee distribution, which can serve as initial information for policymakers in determining the location of coffee development in Bandung Regency. The distribution map of agroforestry and monoculture coffee plants will facilitate policymakers' efforts to assist in the form of seeds, incentives, strengthening production in areas with great potential, and land intervention, as it has been explicitly mapped for coffee commodities. Figure 4 shows some examples of decision tree algorithm rules from a total of 211 branches. Based on the decision tree algorithm, PL Vis is the primary variable separating the classification of coffee plant land cover from others.

The resulting algorithm shows the ability to classify CAF and CNF, with PL Vis = HAL, which serves as the root node of the model. Decision tree rules are used to produce land classification maps in detecting CAF and CNF. The land classification maps formed are natural forests (HAL), plantation forests (HTN), plantations (PKB), rice fields (SWH), water bodies (TBA), open land (TTR), settlements (PMK), CAF, CNF, and dryland agriculture mixed with shrubs (PLKSBL). Figure 4 shows that the decision tree will classify land cover into coffee agroforestry by splitting the elevation branch nodes (DEM/Digital Elevation Models), which have values > 1,513.5 masl, road proximity ≤ 1,373 meters, slope value ≤ 37%, and road proximity > 832 meters. The decision tree will classify coffee monoculture cover by splitting DEM having a value > 1,513.5 masl, distance to the road > 1,373 meters, EVI value > 0.570, MNDWIg value > –0.396, NRGI value ≤ 0.192, ARVI value > –0.999, ARVI value > –0.999, ARVI value > –0.999. The decision tree has many branches to classify coffee agroforestry and coffee monoculture land cover.



Note: ARVI = atmospheric reflection vegetation index, CAF = coffee agroforestry, CNF = coffee monoculture, DEM = elevation, EVI = enhanced vegetation index, HAL = natural forests, HTN = plantation forests, MNDWIg = modified normalized differences wetness index, green base, NRGI = normalized red-green vegetation index, OLC = other land cover, PKB = estate crop plantation, PLKSBL = dryland agriculture mixed with shrubs, plvis = land cover, PROX J = road proximity, Prox_s = river proximity.

Figure 4. Illustrates a decision-tree model used for land-cover classification, showing the hierarchical structure of splitting variables including spectral indices, geophysical parameters, and proximity measures, and the resulting terminal land-cover classes e.g., HAL or natural forest, CAF or coffee agroforestry, and CNF or coffee monoculture. The diagram of decision tree provides context for understanding how different environmental predictors interact and are sequentially selected by the machine-learning algorithm to differentiate between multiple land-cover types. The figure demonstrates that the model relies on a combination of influential spectral and geophysical variables, indicating their importance in accurately discriminating land-cover categories.

Land Cover Classification Results

Following the Indonesian National Standard (SNI/*Standar Nasional Indonesia*) No. 7645–2010 on land cover classification, we developed 10 classes, including coffee agroforestry and monoculture classes, using the best decision tree algorithm obtained. The classification map was then developed and depicted in Figure 5. The algorithm model provides fairly good classification in distinguishing land cover classes. Coffee agroforestry and monoculture are spread across several districts. Coffee agroforestry can be found in 12 districts: Pasir Jambu, Ciwidey, Pangalengan, Cimaung, Banjaran, Rancabali, Pacet, Ibun, Paseh, Cimenyan, Cilengkrang, and Arjasari.

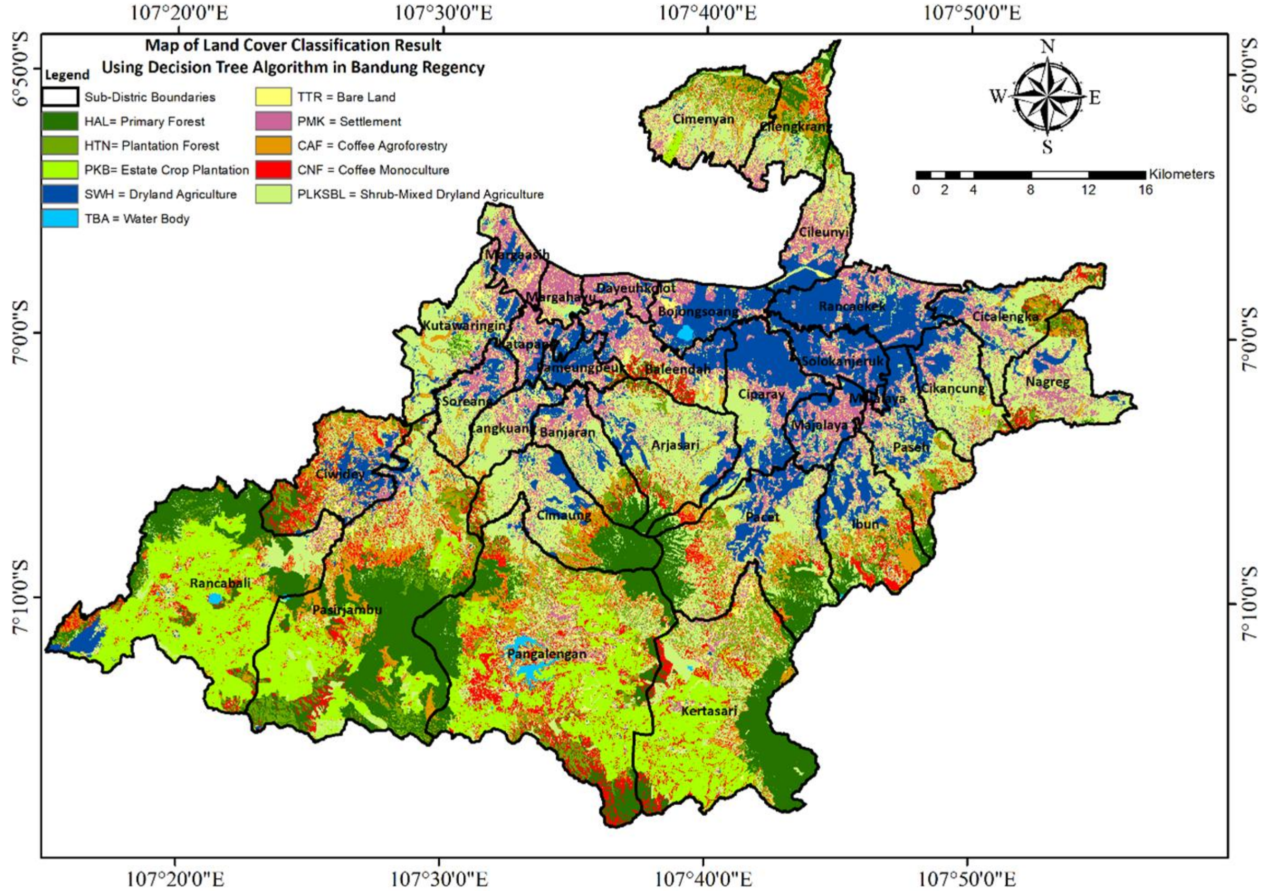


Figure 5. The map displays the spatial distribution of land-cover classes in Bandung Regency, generated using a decision-tree classification algorithm applied to multisource environmental and remote-sensing data. The figure shows eleven distinct land-cover categories, including primary forest, plantation forest, estate crop plantations, coffee agroforestry, coffee monoculture, shrub–mixed dryland agriculture, and settlements, each represented by a unique color. The map highlights substantial heterogeneity in land-cover patterns across the regency, with forested areas concentrated mainly in the southern and southwestern regions, while agricultural and settlement areas dominate the central and northern zones. Coffee agroforestry and coffee monoculture can be found in 12 districts: Pasir Jambu, Ciwidey, Pangalengan, Cimaung, Banjaran, Rancabali, Pacet, Ibun, Paseh, Cimenyan, Cilengkrang, and Arjasari.

In Bandung Regency, coffee agroforestry can be found in highland areas, close to natural and plantation forests. These results are in line with the government program from Perum Perhutani Forest Management Unit (FMU) South Bandung for PHBM [53]. In contrast, coffee monoculture is dominantly found in Baleendah District and the PT Perkebunan Nusantara VIII (PTPN VIII) area due to a shift in commodities planted with tea, replaced by plants without shade. Misclassification still occurs in a set of pixels in several land cover classes. An existing obstacle is overfitting, a weakness of decision trees that occurs with complex datasets. The complexity of the dataset causes the model to follow and adjust the training data [54].

Overfitting might have slightly occurred in Figure 5. One approach to overcoming overfitting is to employ a pruning method. Pruning is the process of removing branches that do not contribute significantly to the model's accuracy value. Complexity is reduced with pruning to generalize new data and reduce overfitting [55]. The classification of coffee monoculture land cover is often wrong due to changes in the function of several areas. The current price increase has led to private agricultural land in the community being converted for coffee plantations. The hedge plants surrounding the monoculture provide a visual resemblance to coffee agroforestry.

User accuracy (UA) provides the percentage of classification results, representing actual conditions in the field [56]. Meanwhile, producer accuracy (PA) provides the percentage of each object in a correctly classified field [57]. UA and PA estimate the overall value, which describes the total value. KA is determined from objects successfully classified correctly and considers classification errors.

Table 9 presents the misclassification of coffee agroforestry cover, including pixels overlapping natural forests, mixed dryland agriculture with shrubs, rice fields, plantation forests, plantations, and coffee monoculture, resulting in a producer accuracy of 68.27%. Coffee agroforestry and coffee monoculture classes are relatively difficult to distinguish in coarse-resolution imagery in a complex landscape. Mixed pixel is always a challenge in image classification. At 30 m resolution, pixels often contain multiple land cover types, complicating classification. This is particularly problematic in heterogeneous landscapes, such as those found in tropical regions. The presence of mixed pixels can lead to lower classification accuracy for specific crops, as the spectral signature of a pixel may represent a combination of coffee and surrounding vegetation [58].

Spectral unmixing techniques, such as spectral unmixing, can help address the mixed pixel problem by decomposing pixel values into their constituent land cover types. This approach has been shown to improve classification accuracy in complex landscapes. The choice of classification algorithm significantly impacts accuracy. For instance, support vector machine (SVM) classifiers have been found to outperform random forest (RF) classifiers at coarser resolutions, suggesting that algorithm choice should be tailored to the specific resolution and landscape complexity [59]. The object-based image analysis (OBIA), which considers both spectral and textural information, has been shown to improve classification accuracy for coffee fields, achieving higher user and producer accuracies compared to pixel-based methods [60]. In this classification, we found that coffee monocultures have a producer accuracy and user's accuracy of 57.3% and 59.7%.

Table 9. The confusion matrix for land-cover classification generated using a combined set of spectral, geophysical, and land-cover variables, showing the number of correctly and incorrectly classified samples for each land-cover category. It provides context for evaluating classification performance through PA and UA, which quantify class-specific reliability from reference data and predicted outcomes, respectively. The table demonstrates strong classification accuracy for most land-cover types particularly water body with UA 93.88%, settlement with UA 93.68%, dryland agriculture with UA 90.64%, and natural forest with UA 89.72%, while highlighting lower performance in classes with higher spectral or structural similarity, such as monoculture coffee.

Prediction class	HAL	CAF	PMK	TBA	TTR	PLKSBL	SWH	HTN	PKB	CNF	**UA (%)**
HAL	96	6	0	0	0	0	0	4	0	1	89.72
CAF	5	71	0	2	0	3	0	11	0	4	73.96
PMK	0	0	178	2	10	0	0	0	0	0	93.68
TBA	0	0	1	46	2	0	0	0	0	0	93.88
TTR	0	0	12	0	47	1	0	0	0	0	78.33
PLKSBL	1	4	1	0	2	165	3	2	1	19	83.33
SWH	0	1	0	6	0	7	155	2	0	0	90.64
HTN	7	17	0	0	0	1	0	79	0	0	75.96
PKB	1	1	0	0	1	2	0	0	113	8	89.68
CNF	0	4	0	0	0	21	0	2	2	43	59.72
PA (%)	87.27	68.27	92.71	82.14	75.81	82.50	98.10	79.00	97.41	57.33	

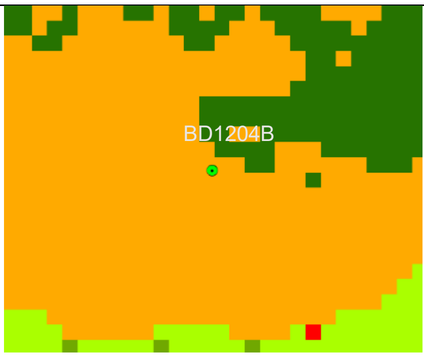
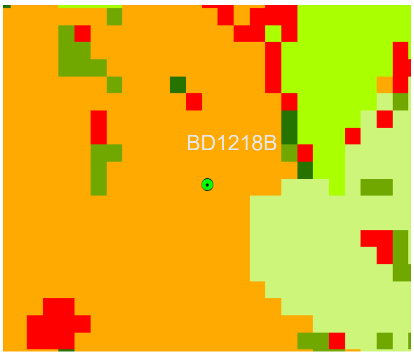
Note: CAF = coffee agroforestry, CNF = monoculture coffee, HAL = natural forest, HTN = plantation forest, PA = producer accuracy, PKB = estate crop plantation, PLKSBL = shrub-mixed dryland agriculture, PMK = settlement, TBA = water body, SWH = dryland agriculture, TTR = bare land, UA = user accuracy.

Table 9 shows that misclassification is not only found in agroforestry and monoculture coffee land cover, but also in other land cover such as natural forest, plantation forest, settlements, water bodies, open land, dry land agriculture, shrubs, and estates due to limitations in visual interpretation. The land cover misclassification occurred due to the limited spatial resolution of the Landsat 8 satellite image used, which only has a spatial resolution of 30 × 30 m. This causes difficulties in interpretation for land cover with similar visualization on the image, for example, open land with settlements, wet rice fields with water bodies, plantation forests with natural forests, and estates with dry land farming shrubs. The integration of spectral indices with geophysical variables has been shown to significantly improve classification accuracy. This research is in line with the results in subtropical forest ecosystems, where the combination of spectral, spatial, and topographic data resulted in an overall classification accuracy of 83.5% for 11 land-cover classes, with specific improvements noted in the classification of coniferous and broadleaf forests. Geophysical variables such as DEM and proximity to roads and rivers provide critical spatial context that enhances the differentiation of land cover types.

For example, the inclusion of topographic factors in spectral and textural imagery improved classification accuracy by up to 5.5% [61]. In mountainous regions, the integration of topographic and geographic information with spectral data has been shown to accurately depict forest distribution and landscape patterns, with an overall accuracy of 95.49% for forest-cover maps [62]. Data fusion techniques, such as the integration of multi-source remote sensing data, have proven effective in enhancing spatial, temporal, and spectral information, leading to improved classification outcomes. For instance, a novel spatial-temporal-spectral fusion framework achieved the highest classification accuracy of 83.6% in distinguishing various forest types [63]. The use of ensemble learning methods and machine learning-based data integration approaches has further improved global-scale forest cover characterization, demonstrating the robustness of these techniques in enhancing classification accuracy [64].

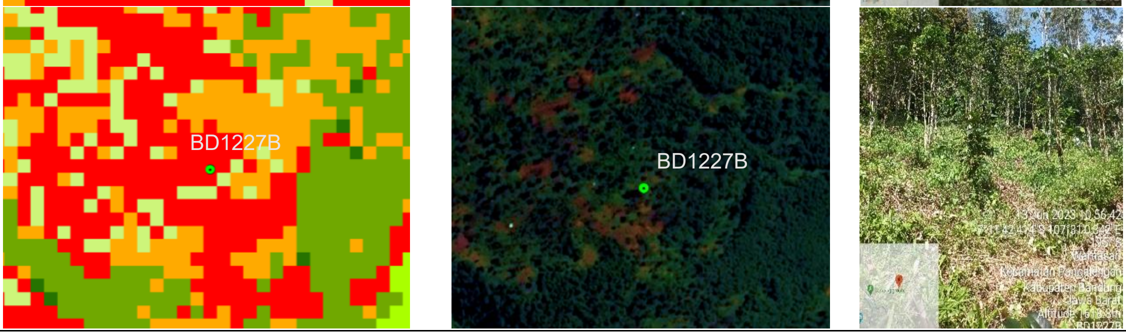
Different factors affect the selection of training areas, resulting in similarity in the pixel values taken. The creation of better training areas and land cover interpretation in Landsat 8 imagery (PL Vis) improves accuracy in classification using a decision tree algorithm [65]. The detection of coffee agroforestry and monoculture coffee resulting from classification using a decision tree algorithm will be compared in appearance through high-resolution imagery and compared with photos taken from field data collection. Comparison of coffee agroforestry and monoculture visualizations can be seen in Table 10.

Table 10. A comparative visualization of CAF and CNF using three complementary data sources: decision-tree classification results, high-resolution satellite imagery, and field-collected ground-truth photographs. The context of the figure is to evaluate how well the decision-tree model differentiates structurally complex agroforestry systems from more uniform monoculture plantations in the landscape. Each row highlights the spatial patterns and visual characteristics of sample locations, showing clear contrasts in canopy structure, vegetation diversity, and surrounding land-cover types. The comparison demonstrates that coffee agroforestry typically appears more heterogeneous and is often situated adjacent to natural forests and mixed agricultural mosaics, whereas monoculture coffee exhibits more homogeneous spectral patterns and is commonly located near plantation forests or cultivated land.

Land cover	Decision tree classification results	High resolution imagery	Ground truth
CAF			
CAF			

CNF

CNF



Note: CAF= coffee agroforestry, CNF= coffee monoculture = natural forest (primary forest), = plantation Forest, = estate crop plantation, = shrub-mixed dryland agriculture, = paddy field, = settlement, = water body, = bare land.

Table 10 shows the proximity of coffee agroforestry and coffee monoculture land cover to others. Coffee agroforestry is adjacent to natural forest land cover, plantations, and dryland shrub farming. This is appropriate to the field conditions because coffee agroforestry is located at high elevations. In this context, natural forests, plantations, and dryland shrub farming are located in the highlands. Coffee monoculture land cover is closely related to plantation forests, coffee agroforestry, rice fields, and open land. In Bandung Regency, coffee monoculture is often found adjacent to community-cultivated land. This is appropriate to the visualization in Figure 4 and Table 10.

Discussion

This study found that the decision tree algorithm was able to classify agroforestry and coffee monoculture land cover correctly. The overall land cover classification accuracy of 84.65% almost resembles the actual conditions in the field. It should be noted that pine trees dominate the characteristics of the coffee agroforestry system in Bandung Regency as shade. This is related to the program of Perum Perhutani that utilizes the pine plantation forest area as a community co-management area with an agroforestry system, so it would be difficult to detect the land cover if relying on visual data of land cover instead. The research findings were that the combination of spectral-geophysical-land cover variables provided the best accuracy in land cover mapping. Spectral variables with vegetation indices such as ARVI, EVI, GARI, NRGI, and VDVI gave a range of –0.99 to 0.57 for coffee monoculture land cover. A vegetation index value range of –0.99 to 0.710 would indicate coffee agroforestry land cover.

Elevation was most influential in separating coffee agroforestry land cover from other land cover in Bandung Regency as a geophysical variable. This is because Arabica coffee is generally planted in the elevation range of 1,500 to 1,750 masl. Higher elevations are associated with improved coffee quality due to cooler temperatures and slower bean maturation, which enhance flavor profiles. This is particularly beneficial in agroforestry systems where shade trees can further modulate microclimates, improving bean quality attributes such as size and biochemical composition [66]. Coffee agroforestry systems at higher elevations are better positioned to adapt to climate change due to their enhanced microclimatic conditions and biodiversity, which can buffer against temperature fluctuations and extreme weather events [67]. Coffee agroforestry systems at higher elevations can experience a trade-off between yield and quality. While the quality may improve, the presence of shade trees can reduce yield compared to monoculture systems. However, the extended maturation period at higher elevations can partially offset this by improving bean size and weight [68,69].

Monoculture systems may achieve higher yields at lower elevations due to full sun exposure, but this often comes at the cost of reduced quality and increased vulnerability to pests and diseases [70]. In monoculture systems, while elevation can still improve quality, the absence of shade trees may lead to less optimal microclimatic conditions, potentially affecting the consistency of quality improvements seen in agroforestry systems [71]. Monoculture systems may struggle with climate adaptation, particularly at lower elevations, where increased temperatures and reduced water availability can exacerbate stress on coffee plants [72]. Coffee monoculture systems, which involve growing coffee in full sun, are often more productive at lower elevations where sunlight is more consistent. Studies have shown that coffee plants in monoculture systems can achieve higher productivity due to increased light exposure, which is beneficial for photosynthesis and growth. At higher elevations, monoculture systems may face challenges, including increased susceptibility to temperature extremes and reduced biodiversity.

These factors can lead to higher pest and disease pressures, necessitating more intensive management practices [73]. Thus, elevation plays a significant role in differentiating between coffee agroforestry and coffee monoculture systems, affecting various aspects, including coffee quality, yield, and environmental sustainability. Coffee agroforestry, which integrates shade trees with coffee plants, often benefits from higher elevations due to cooler temperatures and extended maturation periods, which can enhance coffee quality. In contrast, coffee monoculture, typically grown in full sun, may not leverage these elevation benefits as effectively. The following sections explore how elevation influences these two systems. Slope affects the determination of coffee planting locations chosen by farmers. Farmers have a preference to plant coffee on gentle to steep slopes ≤ 37%.

Coffee plantations do not depend on irrigation, and it is reasonable that the proximity of rivers does not influence coffee planting in Bandung Regency. The proximity of roads provides information on the economic and social linkages of farmers, as coffee will be planted relatively close to roads for easier management, harvesting, and distribution. Other research may provide different results; hence, the findings in this study are only applicable on a local scale and not generalizable to other study areas. Therefore, the key step that can be implemented in other study areas is the method of building the decision tree algorithm, rather than relying on the results of this study as a reference, because they may yield different results. Land cover variables proved unable to provide accurate data on the existence of coffee plantations. Land cover data only provides information on land cover that has proximity to coffee plants, as evidenced by the confusion matrix, which only provides low accuracy for agroforestry and monoculture coffee. Land cover needs to be combined with spectral variables and geophysical variables to classify agroforestry and monoculture coffee using the decision tree method in order to obtain reliable accuracy. It should be underlined that the highest accuracy is not necessarily the best land cover classification model. A model with high accuracy may have overfitting constraints. The optimization of model parameters provides the option of using pruning techniques that can reduce overfitting. The findings of this research prove that pruning can improve accuracy by up to 1%.

This research marks the beginning of a small part of coffee commodity research, which resulted in a map of coffee agroforestry and monoculture distribution in Bandung Regency. The map of coffee agroforestry and monoculture distribution in Bandung Regency can serve as supporting material for decision-makers in formulating policies. After knowing the specific distribution of coffee plantations, the government can formulate policies regarding the development of sustainable coffee plants in the future. Examples of policy formulation using basic data on coffee plant distribution maps are: follow-up to the European Union Deforestation Regulation (EUDR) policy, land traceability, land intervention policy, location of intervention assistance in the form of seeds and incentives, and estimation of coffee productivity throughout Bandung Regency. All of these policies require spatially explicit coffee plants as basic information. In addition, efforts to calculate environmental services also need specific basic data about the area where the value of environmental services will be calculated.

Research on land cover classification with the decision tree method using spectral and geophysical data provides a better accuracy value than the research that has been done, with an overall accuracy of 91.48% [74]. This can occur because the retrieval of training areas in the study is better, and the research area is smaller than Bandung Regency. In line with the research conducted, the use of vegetation Indices and biophysical variables helps distinguish between different land cover types, such as urban areas and vegetation [75]. Additionally, biophysical variables such as DEM and slope are used to improve the separability of coffee agroforestry and coffee monoculture with other classes, such as natural forest, mixed dryland agriculture, settlement, bare land, paddy field, estate crop, and water body [76].

Other studies have examined the performance and accuracy of decision trees, SVM, and Maximum Likelihood Classifiers (MLC). The results consistently show that decision trees provide competitive, even superior accuracy. For example, a study comparing decision trees with SVM and MLC found that decision trees achieved higher accuracy in land cover change assessment. In addition, the integration of decision trees with OBIA has been shown to improve classification accuracy by developing effective rule sets [77]. Decision tree classification, when integrated with Landsat data, vegetation indices, and biophysical variables, provides a powerful tool for land use classification. Its adaptability and accuracy make it a preferred choice in remote sensing applications; however, ongoing research is needed to address existing challenges and further enhance its capabilities. Future research is expected to focus on integrating high-resolution imagery and larger datasets further to improve classification accuracy and applicability in diverse environments [78].

Conclusions

The development of a decision tree algorithm to detect coffee agroforestry and monoculture plantations was successfully carried out using spectral-geophysical-visually based land cover combinations. Spectral-geophysical-land cover combinations variables had an overall accuracy and a kappa accuracy of 84.65% and 82.60%, respectively. Information gain was the criterion selected to detect coffee plantations, both in agroforestry and monoculture systems. The visually-based land cover variable, recognized as the most significant variable, served as the root node in the decision tree. The best parameters used a minimum leaf size set and split of 11 and 31, as well as a maximum depth of 80 through pruning. The coffee agroforestry in Bandung Regency is mainly found in natural forest, having an elevation higher than 1,500 masl, a road proximity of approximately 1 kilometer, a slope of less than 37%, and an ARVI value of less than –0.999. The coffee monoculture class also found close to a natural forest, but outside the natural forest, with elevations of more than 1,500 masl and a road proximity of more than 1 kilometer. The coffee agroforestry and monoculture classes are mainly approached using spectral variables such as ARVI values more or less than –0.999, NRGI values less than 0.192, MNDWIg values more than –0.396, and EVI values more than 0.570. The coffee agroforestry and monoculture were often found close to natural and plantation forests at increasingly higher elevations. Coffee agroforestry is adjacent to natural forest land cover, plantations, and dryland shrub farming. On the other hand, coffee monoculture is often found near community-cultivated land.

Acknowledgments

The authors are grateful to the Economics of Ecosystems and Biodiversity for Agriculture and Food (TEEB Agrifood) Indonesia project, a collaboration between IPB University and the United Nations Environment Programme (UNEP), as well as the National Development Planning Agency (Bappenas), for funding this study. In addition, the authors are grateful to the contributions of reviewers who reviewed the manuscript to improve the quality of the writing. The authors are also grateful to the Division of Remote Sensing and GIS, Department of Forest Management, IPB University, for providing computational facilities.

Authors Contributions

AA: Conceptualization, Data Collection, Data Analysis, Literature Search, Writing Manuscript, Software, Formal Analysis, Resource, Initial Drafting, Design, Editing, Visualization; **INS:** Conceptualization, Critical Review, Methodology, Validation, Investigation, Resources, Editing, Supervision; **NP:** Validation, Writing-Review, Supervision, Editing.

AI Writing Statement

The authors did not use any artificial intelligence assisted technologies in the writing process.

Conflict of Interest

There are no conflicts to declare.

References

1. Agricultural Data and Information Center. *Outlook Komoditas Perkebunan Kopi Tahun 2023*; Directorate General of Agriculture: Jakarta, ID, 2023;
2. BPS (Badan Pusat Statistik). *Statistik Kopi Indonesia Tahun 2021*; BPS: Jakarta, ID, 2021;
3. Pendrill, F.; Gardner, T.A.; Meyfroidt, P.; Persson, U.M.; Adams, J.M.; de Azevedo, T.R.; Lima, M.G.B.; Baumann, M.; Curtis, P.G.; Sy, V.D.; et al. Disentangling the Numbers Behind Agriculture-Driven Tropical Deforestation. *Science* **2022**, *377*, 204–222, doi:https://doi.org/10.1126/science.abm9267.
4. Winara, A.; Fauziyah, E.; Suhartono, S.; Widiyanto, A.; Sanudin, S.; Sudomo, A.; Siarudin, M.; Hani, A.; Indrajaya, Y.; Achmad, B.; et al. Assessing the Productivity and Socioeconomic Feasibility of Cocoyam and Teak Agroforestry for Food Security. *Sustainability* **2022**, *14*, 1–24, doi:doi.org/10.3390/su141911981.
5. Gunawan, B.; Abdoellah, O.S.; Aisharya, I.Y.; Gunawan, W. From Laborers to Coffee Farmers: Collaborative Forest Management in West Java, Indonesia. *Sustainability* **2023**, *15*, 1–22, doi:doi.org/10.3390/ su15097722.
6. Anasrul, A.; Nooraeni, R. Pan-Sharpening Analysis for Improved Detection Accuracy and Estimation of Coffee Plantation Land Area (Case Study: South OKU Regency, South Sumatra Province). *J. Tek. Pertanian Lampung* **2025**, *14*, 424–436, doi:http://dx.doi.org/10.23960/jtep-l.v14i2.424-436.
7. Barus, B.J.A.; Razali, S.; Sitanggang, G. The Evaluation of Land Suitability Coffea Arabica (Coffea arabica L.) in Muara Subdistrict of North Tapanuli District. *Jurnal Online Agroekoteknologi* **2015**, *3*, 1459–1467, doi:10.32734/jaet.v3i4.11797.
8. Nigussie, W.; Al-Najjar, H.; Zhang, W.; Yirsaw, E.; Nega, W.; Zhang, Z.; Kalantar, B. Enhancing Coffee Agroforestry Systems Suitability Using Geospatial Analysis and Sentinel Satellite Data in Gedeo Zone, Ethiopia. *Sensors (Basel)* **2024**, *24*, page range, doi:10.3390/s24196287.
9. Escobar-López, A.; Castillo-Santiago, M.Á.; Mas, J.F.; Hernández-Stefanoni, J.L.; López-Martínez, J.O. Identification of Coffee Agroforestry Systems Using Remote Sensing Data: A Review of Methods and Sensor Data. *Geocarto International* **2024**, *39*, 1–23, doi:10.1080/10106049.2023.2297555.
10. Hehn, T.M.; Kooij, J.F.P.; Hamprecht, F.A. End-to-End Learning of Decision Trees and Forests. *Int. J. Comput. Vis*. **2020**, *128*, 997–1011, doi:10.1007/s11263-019-01237-6.
11. Gaitán-Cremaschi, D.; van Evert, F.K.; Jansen, D.M.; Meuwissen, M.P.M.; Lansink, A.G.J.M.O. Assessing the Sustainability Performance of Coffee Farms in Vietnam: A Social Profit Inefficiency Approach. *Sustainability* **2018**, *10*, 4227–4250, doi:10.3390/su10114227.
12. Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification and Regression Trees*; Wadsworth International Group: Belmont, Massachusetts, USA, 1984;
13. Pal, M.; Mather, P.M. An Assessment of the Effectiveness of Decision Tree Methods for Land Cover Classification. *Remote Sensing of Environment* **2003**, *86*, 554–565.
14. Jaya, I.N.S. *Digital Image Analysis: A Remote Sensing Perspective for Natural Resource Management*, 3rd ed.; IPB University: Bogor, ID, 2015;
15. Amiri, F.; Shariff, A.R.B.M. Using Remote Sensing Data for Vegetation Cover Assessment in Semi-Arid Rangeland of Center Province of Iran. *World Appl. Sci. J*. **2010**, *11*, 1537–1546.
16. Xue, J.; Su, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *J. Sensors*. **2017**, *2017*, 1–17, doi:10.1155/2017/1353691.
17. Gao, B.C.; Li, R.R. FVI—A Floating Vegetation Index Formed with Three Near-IR Channels in the 1.0–1.24 µm Spectral Range for the Detection of Vegetation Floating over Water Surfaces. *Remote. Sens.* **2018**, *10*, 1–10, doi:10.3390/rs10091421.
18. Hu, Y.; Raza, A.; Syed, N.R.; Acharki, S.; Ray, R.L.; Hussain, S.; Dehghanisanij, H.; Zubair, M.; Elbeltagi, A. Land Use/Land Cover Change Detection and NDVI Estimation in Pakistan's Southern Punjab Province. *Sustainability* **2023**, *15*, 1–21, doi:https://doi.org/10.3390/su15043572.
19. Kaufman, Y.J.; Tanré, D. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing* **1992**, *30*, 261–270, doi:https://doi.org/10.1109/36.134076.

20. Justino, S.T.P.; Silva, R.B.; Guerrini, I.A.; Silva, da R.B.G.; Simões, D. Monitoring Environmental Degradation and Spatial Changes in Vegetation and Water Resources in the Brazilian Pantanal. *Sustainability* **2024**, *17*, 51–74, doi:https://doi.org/10.3390/su17010051.
21. Kwan, C.; Gribben, D.; Ayhan, B.; Li, J.; Bernabe, S.; Plaza, A. An Accurate Vegetation and Non-Vegetation Differentiation Approach Based on Land Cover Classification. *Remote Sensing* **2020**, *12*, 1–29, doi:https://doi.org/10.3390/RS12233880.
22. Viña, A.; Henebry, G.M.; Gitelson, A.A. Satellite monitoring of vegetation dynamics: Sensitivity enhancement by the wide dynamic range vegetation index. *Geophysical Research Letters* **2004**, *31*, 1–18, doi:https://doi.org/10.1029/2003GL019034.
23. Selvam, N.; Saranya, R. Analysis of Decision Tree Algorithm in Machine Learning. *Int. J. Adv. Res. Innov. Ideas Educ.* **2018**, *4*, 281–286.
24. Mertes, C.M.; Schneider, A.; Sulla-Menashe, D.; Tatem, A.J.; Tan, B. Detecting change in urban areas at continental scales with MODIS data. *Remote Sensing of Environment* **2015**, *158*, 331–347, doi:10.1016/j.rse.2014.09.023.
25. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: Los Altos, California,USA, 1993;
26. Kumar, A.; Choudhary, T. A Machine Learning Approach for the Land-Type Classification. *Innovations in Electrical and Electronic Engineering*; Mekhilef, S., Favorskaya, M., Pandey, R.K., Shaw, R.N., Eds.; Springer: Singapore, 2021; volume 756, pp. 647–656.
27. Huo, Z.; Martínez-García, M.; Zhang, Y.; Yan, R.; Shu, L. Entropy Measures in Machine Fault Diagnosis: Insights and Applications. *IEEE Transactions on Instrumentation and Measurement* **2020**, *69*, 2607–2620, doi:10.1109/TIM.2020.2981220.
28. Petrović, M.S.; Dragićević, S.; Bajat, B.; Kovačević, M. Exploring The Decision Tree Method for Modelling Urban Land Use Change. *Geomatica* **2015**, *69*, 313–325, doi:10.5623/cig2015-305.
29. Tangirala, S. Evaluating the Impact of Gini Index and Information Gain on Classification using Decision Tree Classifier Algorithm. *International Journal of Advanced Computer Science and Applications* **2020**, *11*, 612–619, doi:10.14569/IJACSA.2020.0110277.
30. Brunello, A.; Marzano, E.; Montanari, A.; Sciavicco, G. Decision Tree Pruning via Multi-objective Evolutionary Computation. *Int J Mach Learn Comput*. **2017**, *7*, 167–175, doi:10.18178/ijmlc.2017.7.6.641.
31. Amro, A.; Al-Akhras, M.; Hindi, K.; Habib, M.; Shawar; B. Instance Reduction for Avoiding Overfitting in Decision Trees. *Journal of Intelligent Systems* **2021**, *30*, 438–459, doi:10.1515/jisys-2020-0061.
32. Mijwil, M.M.; Abttan, R.A. Utilizing the Genetic Algorithm to Pruning the C4.5 Decision Tree Algorithm. *Asian. J. App. Sci*. **2021**, *9*, 45–52.
33. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Morgan Kaufmann: Massachusetts, USA, 2011;
34. Lefkovits, S.; Lefkovits, L. Gabor Feature Selection Based on Information Gain. *Procedia Eng*. **2017**, *181*, 892–898, doi:10.1016/j.proeng.2017.02.482.
35. Rohitha, A. A Comparative Study of the Brute Force Approach with the Hungarian Method of Solving the Travel Distance Problem of Travelling Salesman Problem from Vijayawada to Mangalagiri (Via Tadepalli) to Reach within the Allotted Time. *International Journal of Science and Research* **2023**, *12*, 63–66, doi:10.21275/sr23228200317.
36. Park, Y.; Ho, J. Tackling Overfitting in Boosting for Noisy Healthcare Data. *IEEE Transactions on Knowledge and Data Engineering* **2021**, *33*, 2995–3006, doi:10.1109/TKDE.2019.2959988.
37. Budiarti, W.; Gravitani, E.; Mujiyo. Analysis of Biophysical Aspects for Floods Vulnerability Assessment in Samin Sub-Watershed, Central Java Province. *Journal of Natural Resources and Environmental Management* **2018**, *8*, 96–108, doi:10.29244/jpsl.8.1.96-108.
38. Kumar, U.; Dasgupta, A.; Mukhopadhyay, C.; Ramachandra, T. Examining the Effect of Ancillary and Derived Geographical Data on Improvement of Per-Pixel Classification Accuracy of Different Landscapes. *J Indian Soc Remote Sens*. **2018**, *46*, 407–422, doi:10.1007/s12524-017-0698-2.

39. Dikdayan, G.A.; Ariffin, A. Kajian Iklim Mikro Tanaman Kopi Sistem Agroforestri Di UB Forest. *Produksi Tanaman* **2022**, *10*, 345–349, doi:https://doi.org/10.21776/ub.protan.2022.010.07.01.
40. Andriyani, I.; Ubaidillah, M.M. Penilaian Indikasi Geografis Pegunungan Hyang Argopuro dan Kesesuaian Lahannya untuk Budidaya Kopi. *Agritech*. **2022**, *42*, 131–147, doi:https://doi.org/10.22146/agritech.60195.
41. Sundari, Y.; Asdak, C.; Dwiratna, S. Analisis Karakteristik Fisik Kondisi Lahan di Kabupaten Bandung Barat. In Prosiding Seminar Nasional Pembangunan Dan Pendidikan Vokasi Pertanian, Manokwari, ID, 5 August 2023; pp. 629–635.
42. Chairani, E.; Supriatna, J.; Koestoer, R.H.; Moeliono, M. Physical Land Suitability for Civet Arabica Coffee: Case Study of Bandung and West Bandung Regencies, Indonesia. *IOP Conference Series: Earth and Environmental Science* **2017**, *98*, 1–15, doi:https://doi.org/10.1088/1755-1315/98/1/012029.
43. Ferreira, G.R.; Ferreira, W.P.M.; Barbosa, T.K.M.; Luppi, A.S.L.; Silva, M.A.V. Zoneamento Térmico Para o Cultivo do Café de Montanha na Região das Matas de Minas (Thermal Zoning for Mountain Coffee Crops in the Matas de Minas Region, Brazil). *Revista Brasileira de Geografia Física* **2018,** *11*, 1176–1185, doi:https://doi.org/10.26848/RBGF.V11.4.P1176-1185.
44. Siahaan, A. Identification of Arabica Coffee Production in Altitude Places in Lintong Ni Huta of Humbang Hasundutan. *International Journal of Environment, Agriculture and Biotechnology* **2018**, *1*, 249–255, doi:https://doi.org/10.22161/IJEAB/3.1.31.
45. Somvanshi, S.; Kumari, M. Comparative Analysis of Different Vegetation Indices with Respect to Atmospheric Particulate Pollution Using Sentinel Data. *Applied Computing and Geosciences* **2020**, *7*, 1–10, doi:10.1016/j.acags.2020.100032.
46. Cano, D.; Pizarro, S.; Cacciuttolo, C.; Peñaloza, R.; Yaranga, R.; Gandini, M. Study of Ecosystem Degradation Dynamics in the Peruvian Highlands: Landsat Time-Series Trend Analysis (1985–2022) with ARVI for Different Vegetation Cover Types. *Sustainability* **2023**, *15*, 15472, doi:10.3390/su152115472.
47. Ticehurst, C.; Teng, J.; Sengupta, A. Development of a Multi-Index Method Based on Landsat Reflectance Data to Map Open Water in a Complex Environment. *Remote. Sens.* **2022**, *14*, 1–20, doi:10.3390/rs14051158.
48. Rehman, A.; Ullah, S.; Shafique, M.; Khan, M.; Badshah, M.; Liu, Q. Combining Landsat-8 Spectral Bands with Ancillary Variables for Land Cover Classification in Mountainous Terrains of Northern Pakistan. *Journal of Mountain Science* **2021**, *18*, 2388–2401, doi:10.1007/s11629-020-6548-7.
49. Johnnerie, R.; Siregar, V.P.; Nababan, B.; Prasetyo, L.B.; Wouthuyzen. Random Forest Classification for Mangrove Land Cover Mapping Using Landsat 5 TM and ALOS PALSAR Imageries. *Procedia Environ. Sci.* **2015**, *24*, 215–221, doi:10.1016/j.proenv.2015.03.028.
50. Wardhani, A.K.; Nugraha, E.; Ulfiana, Q. Optimization of the Decision Tree Method using Pruning on Liver Disease Classification. *J. Appl. Informatics Comput*. **2022**, *6*, 136–140.
51. Chern, C.C.; Chen, Y.J.; Hsiao, B. Decision Tree-Based Classifier in Providing Telehealth Service. *BMC Med. Inform. Decis. Mak*. **2019**, *19*, 1–15, doi:10.1186/s12911-019-0825-9.
52. Lopez, A.E.; Santiago, M.A.C.; Mas, J.F. Identification of Coffee Agroforestry Systems using Remote Sensing Data: A Review of Methods and Sensor Data. *Geocarto International,* **2024**, *39*, 1–23, doi:10.1080/10106049.2023.22975.
53. Supriadi, H.; Pranowo, D. Prospects of Agroforestry Development based on Coffee in Indonesia. *Perspektif,* **2015**, *14*, 135–150.
54. Verhaeghe, H.; Nijssen, S.; Pesant, G.; Quimper, C.; Schaus, P. Learning Optimal Decision Trees Using Constraint Programming. *Constraints,* **2020**, *25*, 226–250, doi:10.1007/s10601-020-09312-3.
55. Reiners, M.; Klamroth, K.; Stiglmayr, M. Efficient and Sparse Neural Networks by Pruning Weights in a Multiobjective Learning Approach. *Comput. Oper. Res,* **2022**, *141*, 1–13, doi:10.1016/j.cor.2021.105676.
56. Lee, D.; Kim, H.; Park, J. UAV, a Farm Map, and Machine Learning Technology Convergence Classification Method of a Corn Cultivation Area. *Agronomy* **2021**, *11*, 1–21, doi:10.3390/agronomy11081554.
57. Magnussen, S. Calibration of a Confidence Interval for a Classification Accuracy. *Open Journal of Forestry* **2021,** *11*, 14–36, doi:10.4236/OJF.2021.111002.

58. Löw, F.; Duveiller, G.; Conrad, C.; Michel, U. Impact of Categorical and Spatial Scale on Supervised Crop Classification using Remote Sensing. *Photogrammetrie Fernerkundung Geoinformation*, **2015**, *1*, 7–20, doi:https://doi.org/10.1127/PFG/2015/0252.

59. Awuah, T.K. Effects of Spatial Resolution, Land-Cover Heterogeneity and Different Classification Methods on Accuracy of Land-Cover Mapping. Thesis, Swedish University of Agricultural Sciences, Uppsala, Swedia, 2017.

60. Gaertner, J.; Genovese, V.; Potter, C.; Sewake, K.; Manoukis, N.C. Vegetation classification of Coffea on Hawaii Island using WorldView-2 satellite imagery. *Journal of Applied Remote Sensing,* **2017**, *11*, 1–13, doi:https://doi.org/10.1117/1.JRS.11.046005.

61. Yu, X.; Lu, D.; Jiang, X.; Li, G.; Chen, Y.; Li, D.; Chen, E. Examining the Roles of Spectral, Spatial, and Topographic Features in Improving Land-Cover and Forest Classifications in a Subtropical Region. *Remote Sensing,* **2020**, *12*, 1–24, doi:https://doi.org/10.3390/RS12182907.

62. Wang, Y.; Liu, H.; Sang, L.; Wang, J. Characterizing Forest Cover and Landscape Pattern Using Multi-Source Remote Sensing Data with Ensemble Learning. *Remote Sensing,* **2022**, *14*, 5470–5489, doi:https://doi.org/10.3390/rs14215470.

63. Lu, M.; Chen, B.; Liao, X.; Yue, T.; Yue, H.; Ren, S.; Li, X.; Nie, Z.; Xu, B. Forest Types Classification Based on Multi-Source Data Fusion. *Remote Sensing,* **2017**, *9*, 1153–1175, doi:https://doi.org/10.3390/RS9111153.

64. Ebrahimy, H.; Mirbagheri, B.; Matkan, A.; Azadbakht, M. Effectiveness of the Integration of Data Balancing Techniques and Tree-Based Ensemble Machine Learning Algorithms for Spatially-Explicit Land Cover Accuracy Prediction. *Remote Sensing Applications: Society and Environment,* **2022**, *27*, 2352–2371, doi:10.1016/j.rsase.2022.100785.

65. Zulfajri, D.; Danoedoro, P.; Murti, S.H. Land Cover Classification of Landsat-8 OLI Data using Random Forest Method. *J. Penginderaan Jauh Indonesia* **2021**, *3*, 1–7.

66. Torrez, V.; Benavides-Frias, C.; Jacobi, J.; Ifejika Speranza, C. Ecological Quality as a Coffee Quality Enhancer. A review. *Agronomy for Sustainable Development* **2023**, *43*, 1–34, doi:10.1007/s13593-023-00874-z.

67. Rigal, C.; Vaast, P.; Xu, J. Using Farmers' Local Knowledge of Tree Provision of Ecosystem Services to Strengthen the Emergence of Coffee-Agroforestry Landscapes in Southwest China. *PLOS ONE* **2018**, *13*, e0204046, doi:https://doi.org/10.1371/journal.pone.0204046.

68. Cassamo, C.T.; Draper, D.; Romeiras, M.M.; Marques, I.; Chiulele, R.; Rodrigues, M.; Stalmans, M.; Partelli, F.L.; Ribeiro-Barros, A.; Ramalho, J.C. Impact of Climate Changes in the Suitable Areas *for Coffea arabica L.* Production in Mozambique: Agroforestry as an Alternative Management System to Strengthen Crop Sustainability*. Agriculture, Ecosystems and Environment* **2023**, *346*, 108341, doi:https://doi.org/10.1016/j.agee.2022.108341.

69. Koutouleas, A.; Sarzynski, T.; Bordeaux, M.; Bosselmann, A. S.; Campa, C.; Etienne, H.; Turreira-García, N.; Rigal, C.; Vaast, P.; Ramalho, J. C.; et al. Shaded-Coffee: A Nature-Based Strategy for Coffee Production Under Climate Change? A Review. *Frontiers in Sustainable Food Systems* **2022**, *6*, 1–21, doi:10.3389/fsufs.2022.877476.

70. Acosta-Alba, I.; Boissy, J.; Chia, E.; Andrieu, N.; Andrieu, N. Integrating Diversity of Smallholder Coffee Cropping Systems in Environmental Analysis. *International Journal of Life Cycle Assessment* **2020**, *25*, 252–266, doi:https://doi.org/10.1007/S11367-019-01689-5.

71. Bertrand, B.; Vaast, P.; Alpizar, E.; Etienne, H.; Davrieux, F.; Charmetant, P. Comparison of Bean Biochemical Composition and Beverage Quality of Arabica Hybrids Involving Sudanese-Ethiopian Origins with Traditional Varieties at Various Elevations in Central America. *Tree Physiology,* **2006**, *26*, 1239–1248, doi:https://doi.org/10.1093/TREEPHYS/26.9.1239.

72. Mayorga, I.; Mendonça, J.L.V.de; Hajian-Forooshani, Z.; Lugo-Pérez, J.; Perfecto, I. Tradeoffs and Synergies among Ecosystem Services, Biodiversity Conservation, and Food Production in Coffee Agroforestry. *Frontiers in Forests and Global Change* **2022**, *5*, 1–15.

73. Farfán-Valencia, F. Sistemas Agroforestales con Café: Establezca Cultivos Productivos Bajo Sombrío. *Memorias Seminario Científico Cenicafé* **2022**, *71*, 63–78.

74. Prayogo, L.; Widyantoro, B.; Yuliardi, A.; Hanif, M.; Spanton, P.; Joesidawati, M. Land Cover Classification Assessment Using Decision Trees and Maximum Likelihood Classification Algorithms on Landsat 8 Data. *Journal of Computer and Information Technology* **2023**, *6*, 69–76, doi:10.25273/doubleclick.v6i2.10606.
75. Hua, L.; Zhang, X.; Chen, X.; Yin, K.; Tang, L. A Feature-Based Approach of Decision Tree Classification to Map Time Series Urban Land Use and Land Cover with Landsat 5 TM and Landsat 8 OLI in a Coastal City, China. *Int. J. Geo Inf.* **2017**, *6*, 331–349, doi:10.3390/ijgi6110331.
76. Yang, C.; Wu, G.; Ding, K.; Shi, T.; Li, Q.; Wang, J. Improving Land Use/Land Cover Classification by Integrating Pixel Unmixing and Decision Tree Methods. *Remote. Sens*. **2017**, *9*, 1222–1238, doi:10.3390/rs9121222.
77. Phiri, D.; Simwanda, M.; Nyirenda, V.; Murayama, Y.; Ranagalage, M. Decision Tree Algorithms for Developing Rulesets for Object-Based Land Cover Classification. *Int. J. Geo Inf*. **2020**, 9, 329–337, doi:10.3390/ijgi9050329.
78. Hailu, B.T.; Maeda, E.E.; Pellikka, P.; Pfeifer, M. Identifying Potential Areas of Understorey Coffee in Ethiopia's Highlands Using Predictive Modelling. *Int J Remote Sens*. **2015**, *36*, 2898–2919, doi:10.1080/01431161.2015.1051631.