



Article

Critical Assessment of Cocoa Classification with Limited Reference Data: A Study in Côte d'Ivoire and Ghana Using Sentinel-2 and Random Forest Model

Nikoletta Moraiti ^{1,*} , Aduigna Mullissa ^{1,2,3}, Eric Rahn ⁴, Marieke Sassen ⁵ and Johannes Reiche ¹

¹ Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Droevendaalsesteeg 4, 6708 PB Wageningen, The Netherlands; amullissa@ucla.edu (A.M.); johannes.reiche@wur.nl (J.R.)

² Ctrees.org, 12 S Raymond Avenue, Pasadena, CA 91105, USA

³ Institute of the Environment and Sustainability, University of California Los Angeles, 619 Charles E. Young Drive East, Los Angeles, CA 90095, USA

⁴ International Center for Tropical Agriculture (CIAT), Km 17 Recta Cali-Palmira, Cali 763537, Colombia; e.rahn@cgiar.org

⁵ Plant Production Systems Group, Wageningen University, Bornsesteeg 48, 6708 PE Wageningen, The Netherlands; marieke.sassen@wur.nl

* Correspondence: moraiti.nikol@gmail.com

Abstract: Cocoa is the economic backbone of Côte d'Ivoire and Ghana, making them the leading cocoa-producing countries in the world. However, cocoa farming has been a major driver of deforestation and landscape degradation in West Africa. Various stakeholders are striving for a zero-deforestation cocoa sector by implementing sustainable farming strategies and a more transparent supply chain. In the context of tracking cocoa sources and contributing to cocoa-driven deforestation monitoring, the demand for accurate and up-to-date maps of cocoa plantations is increasing. Yet, access to limited reference data and imperfect data quality can impose challenges in producing reliable maps. This study classified full-sun-cocoa-growing areas using limited reference data relative to the large and heterogeneous study areas in Côte d'Ivoire and Ghana. A Sentinel-2 composite image of 2021 was generated to train a random forest model. We undertook reference data refinement, selection of the most important handcrafted features and data sampling to ensure spatial independence. After refining the quality of the reference data and despite their size reduction, the random forest performance was improved, achieving an overall accuracy of $85.1 \pm 2.0\%$ and an F1 score of $84.6 \pm 2.4\%$ (mean \pm one standard deviation from ten bootstrapping iterations). Emphasis was given to the qualitative visual assessment of the map using very high-resolution images, which revealed cases of strong and weak generalisation capacity of the random forest. Further insight was gained from the comparative analysis of our map with two previous cocoa classification studies. Implications of the use of cocoa maps for reporting were discussed.

Keywords: cocoa classification; crop mapping; limited reference data; qualitative assessment; random forest; feature engineering; Sentinel-2; Ghana; Côte d'Ivoire



Citation: Moraiti, N.; Mullissa, A.; Rahn, E.; Sassen, M.; Reiche, J. Critical Assessment of Cocoa Classification with Limited Reference Data: A Study in Côte d'Ivoire and Ghana Using Sentinel-2 and Random Forest Model. *Remote Sens.* **2024**, *16*, 598. <https://doi.org/10.3390/rs16030598>

Academic Editor: Eufemia Tarantino

Received: 28 November 2023

Revised: 25 January 2024

Accepted: 31 January 2024

Published: 5 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Land-use and land-cover maps generated from remotely sensed data are valuable sources of information for the state of Earth's surface and its changes over time [1]. These mapping products serve as particularly useful tools for applications that study and monitor biodiversity, sustainable land management and forest conservation, to mention a few examples [2–4]. Land-cover information also plays a key role in setting and monitoring progress towards several of the United Nations' Sustainable Development Goals [5].

Over the last decade, land-use/cover mapping has been greatly facilitated by a plethora of publicly available remote sensing imagery [6–8] and advancements in computational processing capacity and data storage [9]. Thanks to the viability and scalability of

these satellite images and their high spatial and temporal resolution, researchers have been enabled to carry out challenging global mapping studies [10,11], as opposed to the use of proprietary images, which are characterised by scale limitations and high cost.

Good-quality land-use/cover maps are required to deliver reliable outputs to support policy decisions and prioritise actions, especially when the maps are linked to other datasets or models [12,13]. Ensuring the good quality of a classification product is, however, a complex task and depends on various factors that can have a considerable impact on map accuracy. These factors involve, among others, the reference data size and their quality [14,15], the sampling design [16], and the classification model and its parameterisation [17,18].

Supervised machine learning models require sufficient and representative training and test data to perform effectively and prevent overfitting [19,20]. However, acquiring adequately labeled data of good quality remains a big challenge in the field of remote sensing [5]. Reference datasets are typically collected with costly and labour-intensive field surveys and/or by interpreting greater resolution images than those used in classification. Practical reasons, such as a limited budget, time constraints and inaccessible areas, may limit field campaigns at the spatial and temporal levels. Photointerpretation, on the other hand, has considerably lower cost and can sometimes be equally effective to field visits [21], but it requires expertise in land-type interpretation [22]. Visual interpretation can be additionally constrained in chronically cloudy areas, e.g., in the tropics during the wet season. Nevertheless, the gradual outdateding of the reference data, as well as the high demand for releasing frequent classification maps, renders data collection a highly demanding task [5].

Although reference data are expected to be more accurate than the classification outcome [23], they can be subject to errors regardless of the collection method used. For instance, inadequately trained field staff or farmers' self-reporting might result in unreliable measurements [24,25], whilst photointerpreted datasets can be associated with uncertainty [22], and some degree of disagreement is anticipated among multiple experts [26]. The primary error sources that downgrade the quality of reference data can be thematic mis-labeling, position mis-registration and mis-synchronisation with the remote sensing imagery used [27,28]. Allocating correct labels can be particularly confounded when data collection takes place in thematically heterogeneous areas (i.e., where mixed pixels dominate) [26], and when different land types look spectrally similar [29]. Either random or systematic errors in reference data can adversely influence an accuracy assessment [29,30] and ultimately result in area mis-estimation [31]. However, correcting erroneous reference datasets was shown to improve map accuracy [32].

High-quality crop maps are currently the subject of focus for detecting the location and estimating the yields of commodity tree crops, like cocoa, coffee and oil palm [33–35]. This is because the cultivation of such tropical cash crops, which is integral to the global supply chain and a livelihood source for millions of farmers, has considerable environmental impacts [36]. In particular, cocoa farming has attracted the spotlight for being a major deforestation factor and driver in the degradation of protected areas in West Africa [37,38]. In response to the high deforestation risks, intervention programmes have been put in place to help cocoa-producing countries and companies halt cocoa-driven deforestation and restore degraded areas. The Cocoa and Forests Initiative (CFI) is such a programme where Côte d'Ivoire and Ghana, which are the biggest cocoa producers worldwide, and industry stakeholders are committed to a deforestation-free and sustainable cocoa sector [39]. Mondelez International and Nestlé, who are members of CFI, have set action plans for mapping and monitoring their supply chain at the farm level with regard to forest protection [40,41]. Moreover, the European Union has adopted a new regulation to ensure that imported commodities, including cocoa, are not associated with deforestation and forest degradation [42]. Mapping cocoa expansion in an accurate and timely way can therefore be a valuable tool for assessing cocoa-driven deforestation and monitoring landscape restoration efforts through agroforestry [34,43,44].

Over the past decade, a few studies on cocoa mapping have been conducted using wall-to-wall remote sensing data in West and Central Africa [34,43,45–48]. A commonality we

observed between most of them is the use of limited reference data, where the total magnitude of the reference samples ranged from 1838 to 6450 pixels at the landscape scale [43,48], and 19,196 pixels over a larger area [45] at a 10×10 m pixel resolution. Some of these studies called their mapping techniques “effective” or “successful” based solely on confusion matrix-derived accuracy measures and the general look of their thematic map [43,45,47,48]. It was often neglected to reflect on both the strengths and the shortcomings of the classification outcome or provide insight into the potential impacts of the limited reference dataset. Yet, it is of great importance to convey map accuracy in an informed way and to communicate its potential or likely errors and their implications for the intended purposes, particularly when classified maps constitute the basis for further analysis [28].

The objective of our study was to explore the potential of classifying cocoa growing areas with the use of limited reference data over a large and heterogeneous study area in Côte d’Ivoire and Ghana. We undertook quality refinement of our limited reference data, sampling design and feature selection for training a random forest model using Sentinel-2 imagery. To assess our cocoa map, we performed an accuracy assessment with quantitative measures and a visual map inspection to highlight classification strengths and errors. We further compared our results with two previous cocoa studies and discuss the implications for reporting on cocoa area estimation.

2. Materials and Methods

2.1. Study Area

We delineated our study area with a bounding box determined by the available reference cocoa data located in Côte d’Ivoire and Ghana (Figure 1a). A total surface area of approximately 17.4 Mha was covered, which corresponded to the most suitable areas for cocoa cultivation [49,50]. The study area fell within latitudes $5^{\circ}5'58.87''\text{N}$ – $7^{\circ}19'16.63''\text{N}$ and longitudes $0^{\circ}18'31.18''\text{W}$ – $6^{\circ}41'35.13''\text{W}$. The Ghanaian regions of Western, Western North, Ashanti, Ahafo, Central, Eastern, and small parts of Bono East and Greater Accra were incorporated into the study area. Accordingly, the Ivorian districts of Comoé, Lagunes, Abidjan, Lacs, Yamoussoukro, Sassandra-Marahoué, Gôh-Djiboua and Bas-Sassandra were included (Figure 1b).

Côte d’Ivoire has four large agro-climatic zones [51], with our study area extent falling within the forest zone and Guinea savannah. There, the climate is wet and tropical, and the annual rainfall ranges from 1000 to 1600 mm [51]. The southern and central part of Côte d’Ivoire experiences two rainy and two dry seasons. The topography is defined by a large plateau featuring undulating plains and mountains in the north-western part.

Ghana is divided into five agro-ecological zones [52]. Our study area fell mainly under the rainforest zone, deciduous forest zone, transitional zone and coastal savannah. In these zones, the average rainfall ranges from 900 to 1800 mm annually, resulting in a major and a minor growing season [52]. The climate is hot and humid in the southern and western parts, and warm and dry in the coastal zone [52]. Ghana’s topography is primarily undulating with mild slopes [52].

Lowland forests in Côte d’Ivoire and Ghana are part of the Guinean forests of West Africa, which is a globally important biodiversity hotspot [53]. Agriculture managed by smallholder farmers is the dominant land-use in both countries [51,52]. Their cropping systems mainly include perennial cash crops (e.g., cocoa, cashew, rubber and palm oil) and food crops (e.g., rice, maize, cassava and yam) [51,52]. Agricultural activities, especially growing cocoa, have historically been to the detriment of the forests [38,54,55], creating a heterogeneous landscape composed of a mosaic of annual and perennial croplands and forest reserves, which are characteristic in Côte d’Ivoire and Ghana [37,46,52,53].

2.2. Reference Data

2.2.1. Raw and Refined Cocoa Polygons and Non-Cocoa Polygons Curation

The boundaries and centre locations of cocoa plot polygons were collected with in situ global positioning system (GPS) measurements in Côte d'Ivoire and Ghana from December 2019 until August 2020. In total, 93 cocoa polygons of varying sizes were collected with an average area of 2.3 hectares. These plots mainly included full-sun cocoa trees mixed with shade trees in varying canopy levels, while additional land-use/cover types were recorded in some cases, such as bare ground and unpaved roads (see example plots in Figure 1c). We hereafter refer to these polygons as raw cocoa to represent the lack of quality refinement.

To improve the thematic homogeneity in the raw cocoa polygons and the purity in the cocoa spectral responses, we refined the raw cocoa by manually digitising polygons based on pixels that included full-sun cocoa trees, or the least shaded cocoa in cases where we could not clearly distinguish cocoa from shade trees (Figure 1c). The refined full-sun cocoa tree pixels could be part of monocultures and/or agroforestry systems. The fact that full-sun cocoa grows in self-standing and unshaded plantations renders them visually distinct for photointerpretation. In contrast, cocoa trees in agroforestry systems are mostly planted in the understorey of forest trees and/or other tree crops. As a result, they look very similar to the natural forest, especially when matured, and they are not visually distinguishable [56]. We hereafter refer to the refined cocoa polygons as refined cocoa.

To ensure the quality of our cocoa refinement method, we visually interpreted very high-spatial-resolution Google Earth images dated between 2018 and 2022. Due to the low image availability caused by the persistent cloud coverage in West Africa, we decided to photointerpret Google Earth images captured within a longer time span. Although these capture dates were not perfectly aligned with the time frame of the field measurements, we considered that this deviation would not have influenced the refinement method because cocoa is a perennial crop and can remain planted for more than 25 years [57,58].

To distinguish cocoa from other land-use/cover types, we conducted a binary classification. An inventory of true negative samples was manually labelled by photointerpreting Google Earth images captured between 2018 and 2022. The curated dataset included 3311 polygons, which represented diverse land-use/cover types across different locations within the study area. These polygons, referred to as non-cocoa hereafter, included forest, oil palm, rubber, shade trees in agroforestry systems, shrub vegetation, bare soil, grassland, built-up areas, water bodies, and unpaved and paved roads. The complete dataset of the refined cocoa and the non-cocoa polygons is hereafter referred to as the refined dataset; conversely, the term raw dataset is used.

2.2.2. Sampling Design

We split our reference data into three distinct datasets for training, validation and testing. The training set was used for the learning process of the classification model, while the validation set was used for estimating the importance of handcrafted features and for tuning the model's hyperparameters. The test set was used to assess the classification accuracy. To ensure independence between the training, validation and test datasets, we designed a splitting scheme that accounted for spatial autocorrelation. The issue of spatial autocorrelation can be caused by the similarities that spatially adjacent observations tend to have, which leads to inflated accuracy and overestimated model predictive power [59,60]. Therefore, drawing random samples from the entire spatial domain would not have guaranteed independence between datasets.

Our scheme for splitting the data into training, validation and test datasets was designed separately for cocoa and non-cocoa plots due to their disparate distribution across the study area. Since our acquired cocoa plots were distributed unevenly, we did not split them at the polygon level because groups of polygons were located close to each other (Figure 1b). Instead, we split them into spatially homogeneous groups with a minimum inter-group distance of 50 km, i.e., training, validation and test cocoa datasets were at least

50 km apart from each other. In addition to that, we took care that each dataset represented different locations across the study area (Figure 1b).

The distribution of the raw and refined cocoa datasets was identical, as both were derived from the same cocoa polygons, but the number of cocoa pixels dropped to approximately half after applying refinement. Our splitting scheme resulted in 68 cocoa polygons for training, from which 9427 raw and 4437 refined cocoa pixels were derived; 14 cocoa polygons for validation, from which 405 raw and 186 refined cocoa pixels were derived; and 11 polygons for testing, from which 723 refined cocoa pixels were derived (see Table 1). The refined test dataset was used for assessing all classification schemes to ensure a fair comparison. The number of pixels refers to a 10×10 m resolution.

Table 1. Number of cocoa pixels and number of cocoa polygons they were derived from, and number of non-cocoa pixels as split for the training, validation and test datasets before and after the cocoa refinement. For the accuracy assessment of all classification schemes, only the refined reference dataset was used. Pixel size refers to 10×10 m spatial resolution.

Data	Reference Dataset	Cocoa Pixels	Cocoa Polygons	Non-Cocoa Pixels
Training	Raw	9427	68	9322
	Refined	4437		4392
Validation	Raw	405	14	405
	Refined	186		186
Test	Refined	723	11	719

Curated non-cocoa polygons were distributed more evenly across the study area compared with the cocoa dataset (see Figure A1). For this reason, we implemented a gridded block sampling design [60]. In particular, we created a grid of irregular blocks where each one incorporated a relatively consistent number of non-cocoa polygons. Spatially disjoint blocks were used to draw samples for the training, validation and test datasets to ensure diverse regional coverage for each dataset. From the respective blocks, samples were drawn randomly, yet in such proportions to achieve a total size approximately equal to that of the cocoa dataset (Table 1).

2.3. Sentinel-2 Imagery Pre-Processing

Sentinel-2 level-2A images were the input data to our classification and consisted of twelve spectral bands with varied spatial resolutions of 10 m, 20 m and 60 m, and a revisit time of five days with the twin satellite constellation [7]. Sentinel-2 level-2A products are atmospherically corrected orthoimages and correspond to bottom-of-atmosphere reflectance [61]. The Sentinel-2 archive was accessed via the Google Earth Engine's (GEE) data catalogue and processed in GEE's cloud-computing environment [9].

We defined 2021 as our study period. A median composite image was generated from all Sentinel-2 level-2A images captured in 2021. The median method was deemed the most appropriate because of its ability to disregard reflectance outliers and its smoother output compared with other compositing methods [62]. We considered the one-year time span as a sufficient period for capturing the representative temporal dynamics of cocoa trees and for generating a quality cloud-free composite given that our study area was susceptible to high cloud coverage.

To remove clouds, we applied the s2cloudless algorithm, which has been developed for Sentinel-2 imagery and detects clouds per pixel at a global scale based on machine learning [63]. Its output is a cloud probability map for a given scene and it is provided pre-computed in GEE as a separate yet compatible image collection [64]. A recently published comparison study evaluated s2cloudless as one of the best-performing cloud detection algorithms in terms of the user's and producer's accuracy [65]. The s2cloudless parameters are user-defined [66]. After testing different parameter values, we opted for the ones that

generated fewer cloud artefacts (Table A1). Specifically, we set the maximum cloud coverage at 100% to consider all image scenes captured in 2021 for the compositing method. This way, 0.5% of our total reference data were masked out, whereas cloud coverage percentages lower than 100% yielded a larger number of masked pixels. We followed an aggressive cloud masking to ensure the removal of thin cloud residuals. Hence, we set the cloud probability threshold at 20%, above which a pixel was identified as cloudy. The s2cloudless algorithm also detects cloud shadows by intersecting detected clouds with low-reflectance near infrared pixels [66]. In our study, pixels with less than 0.2 near-infrared reflectance were considered cloud shadows, which were examined within a 1.5 km distance from cloud edges, and a buffer of 50 m was determined around cloud-identified objects. The cloud and cloud shadow masking was implemented on each image scene from the 2021 Sentinel-2 level-2A collection. Following that, the median compositing method was applied. The resultant 2021 median composite is illustrated in Figure A2.

2.4. Feature Engineering

Feature engineering is an essential step in classification tasks, especially for conventional machine learning models. We derived handcrafted features from the 2021 median composite image (i.e., in our study the term feature was used to refer to a derived variable predictor). When selecting the types of features to be generated, we drew inspiration from previous cocoa studies [43,45,48,67] and other land-use/cover classification studies [68–71]. We initially generated a series of candidate features that served as a preliminary pool on which we applied the feature selection. The finalised subset of features was determined by their relative importance ranking, as derived from the permutation-based importance measure. Our handcrafted features were divided into two main categories: vegetation indices (Table 2) and textural features (Table 3).

2.4.1. Vegetation Indices

A total of eleven vegetation indices (VIs) were generated, which served as candidate features in the feature selection method. Table 2 provides a comprehensive description of the derived VIs with the corresponding formulas. The list included normalised difference indices, ratio indices and triangular indices. We derived VIs that combined near-infrared and red wavelengths to strengthen the presence of photosynthetically active vegetation; VIs that leveraged green and red edge spectra, which are more sensitive to higher chlorophyll concentration [72]; and VIs that exploited the shortwave infrared spectrum to highlight vegetation moisture. Moreover, we calculated VIs with less sensitivity to soil background and atmospheric effects.

2.4.2. Textural Features

To generate textural features, we employed the grey-level co-occurrence matrix (GLCM), which is the most applied method for quantifying textures at the pixel level. A GLCM is in essence a matrix that describes the number of occurrences for pair-wise pixel values that occur in the original image expressed as probabilities [73]. There are various textural measures to summarise a GLCM, which were originally proposed by [74]. In this study, we calculated 17 GLCM-based textural measures for the 12 Sentinel-2 bands (Table 3), resulting in 204 textural features in total. The respective built-in function in GEE was used for this purpose [9]. Since the GLCM implementation was only supported with integer values, we converted reflectance values into 32-bit integers. A 3×3 -pixel moving window was applied, averaged over four spatial directions (i.e., horizontal, vertical, two diagonals) with a one-pixel displacement. This window size defined the neighbourhood where texture calculations took place, assigning the output to the window's centre pixel. We opted for a small window size with the aim to capture the texture of small-sized cocoa plots, following the practice of [45].

Table 2. The 12 reflectance bands of the Sentinel-2 median composite image, along with 11 derived vegetation indices (VIs). Formulas were adjusted to the Sentinel-2 bands. Different names for VIs may have been used in literature.

Input Data Type	Description	Acronym	Formula	Source
Sentinel-2 level-2A reflectance bands	Coastal aerosol	B1	-	-
	Blue	B2	-	-
	Green	B3	-	-
	Red	B4	-	-
	Red-edge 1	B5	-	-
	Red-edge 2	B6	-	-
	Red-edge 3	B7	-	-
	Near-infrared	B8	-	-
	Red-edge 4	B8A	-	-
	Water vapor	B9	-	-
	Shortwave infrared 1	B11	-	-
	Shortwave infrared 2	B12	-	-
Vegetation indices	Enhanced vegetation index	EVI	$\frac{2.5 \times (B8 - B4)}{B8 + 6 \times B4 - 7.5 \times B2 + 1}$	[75]
	Green normalised difference vegetation index	GNDVI	$\frac{B8 - B3}{B8 + B3}$	[76]
	Normalised difference moisture index	NDMI	$\frac{B8 - B11}{B8 + B11}$	[77]
	Normalised difference red edge index	NDRE	$\frac{B8 - B5}{B8 + B5}$	[78,79]
	Normalised difference vegetation index	NDVI	$\frac{B8 - B4}{B8 + B4}$	[80]
	Plant senescence reflectance index	PSRI	$\frac{B4 - B2}{B6}$	[81]
	Red-edge chlorophyll index	RECI	$\frac{B7}{B5} - 1$	[72,82]
	Red-edge normalised difference vegetation index	RENDVI	$\frac{B6 - B5}{B6 + B5}$	[83]
	Soil-adjusted vegetation index	SAVI	$\frac{B8 - B4}{B8 + B4 + 0.428} \times 1.428$	[84]
	Triangular chlorophyll index	TCI	$1.2 \times (B5 - B3) - 1.5 \times (B4 - B3) \times \text{SQRT}(\frac{B5}{B4})$	[85]
	Visible atmospherically resistant index	VARI	$\frac{B3 - B4}{B3 + B4 - B2}$	[86]

Table 3. Seventeen textural measures based on the grey-level co-occurrence matrix (GLCM) were computed for the 12 Sentinel-2 bands. Their abbreviations are within parentheses.

GLCM Textural Measures		Source
Angular second moment (asm)	Inertia (inertia)	[74,87]
Cluster prominence (prom)	Information measure of correlation 1 (imcorr1)	
Cluster shade (shade)	Information measure of correlation 2 (imcorr2)	
Contrast (contrast)	Inverse difference moment (idm)	
Correlation (corr)	Sum average (savg)	
Difference entropy (dent)	Sum entropy (sen)	
Difference variance (dvar)	Sum variance (svar)	
Dissimilarity (diss)	Variance (var)	
Entropy (ent)		

Given that some textural features showed a strong correlation with each other due to their similar calculation method [73], it is suggested to select a subset of them [74]. We implemented agglomerative hierarchical clustering to select a less-correlated subset from the 204 generated textural features. To this end, we employed the Scikit-learn package [88] in the Google Colaboratory environment [89]. Agglomerative hierarchical clustering progressively groups similar observations into clusters, starting by treating single observations

as separate clusters [90]. To determine the clusters' pairing, the concept of linkage was computed on the chosen similarity metric. We calculated the Spearman rank-order correlation matrix for the training data and applied the Ward linkage method, which minimises the total within-cluster variance [91]. The raw dataset and the refined dataset were treated separately, yielding two dendrograms (see Figure A3). After visual inspection, we chose the cut-off value at 1.5 for both dendrograms since clusters below this threshold were densely grouped, indicating a higher correlation. The cutting threshold yielded 14 clusters in both datasets. From each cluster, one feature was randomly chosen, resulting in 14 textural features (see highlighted names in Figure A3), which served as candidate features in the feature selection method that followed.

2.4.3. Feature Selection

Feature selection is key to the performance of supervised machine learning models. Reducing feature space dimensions and removing redundant or irrelevant features can facilitate the model generalisation capacity and can result in a lower computational cost [92,93]. Feature importance measures can serve the purposes of feature selection [94]. In this study, we selected a subset from the 11 VIs and 14 textural features based on their relative importance with respect to their relevance to discriminate cocoa from the other land types. For this purpose, the permutation-based feature importance measure was employed (hereafter referred to as permutation), which was initially introduced for the random forest by [95]. We opted for permutation instead of the commonly used Gini index because permutation can be applied to any dataset whilst Gini is performed exclusively on the training dataset [95]. In our study, we applied permutation on the held-out validation data of the raw dataset and the refined dataset, after training the random forest on the respective training samples.

Permutation measures the accuracy when the values of a particular feature are randomly permuted and computes the difference in accuracy before and after the permutation [95]. A decrease in accuracy indicates that the model's performance relies on that particular feature. The rationale behind this concept is that the relationship of the permuted feature with the true outcome breaks, as well as with the remaining features [94,96]. However, it has been reported that correlation between features can have a great impact on permutation [94,96], as the actual importance may be diffused when correlated features hold similar information. We ensured to eliminate the strongly correlated textural features by applying hierarchical clustering in the previous step.

We estimated the importance of the 11 VIs and 14 textural features separately since they relate to different types of measurement. Permutation was repeated 20 times for each feature and the decrease in accuracy was averaged. Based on the mean decrease in accuracy, the features were ranked in decreasing importance order for the raw dataset (Figure 2a,b) and the refined dataset (Figure 2c,d), respectively. We performed permutation using the Scikit-learn package [88] in Google Colaboratory [89].

The interpretation of the output was straightforward, e.g., NDVI yielded a 13% mean decrease in accuracy when it was randomly permuted and it was ranked first in importance for the refined dataset (Figure 2c). Ultimately, for the final selection, we considered the features that yielded a considerable decrease in accuracy. We selected the VIs with a mean decrease in accuracy greater than 8%, and the textural features with a mean decrease in accuracy greater than 6%. We ended up selecting the following:

- TCI, NDVI, GNDVI and B6_savg for the raw dataset;
- NDVI, TCI, B12_savg and B8_savg for the refined dataset.

The features that did not yield a decrease in accuracy were not taken into consideration. This was an indication that the feature set with the permuted feature happened to achieve higher accuracy than the original feature set, which was caused by random chance [97].

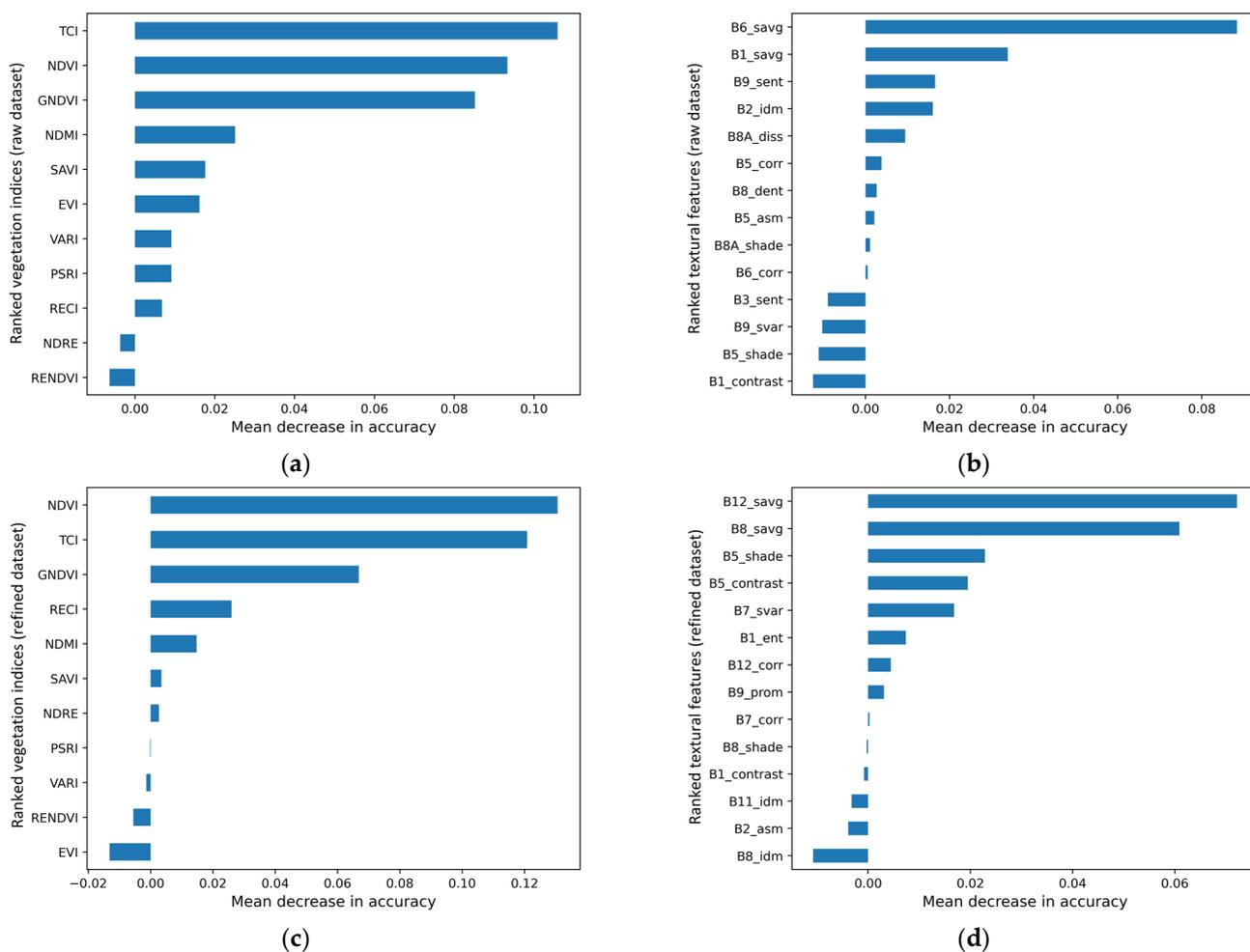


Figure 2. The feature importance rankings as estimated by the permutation-based measure for the vegetation indices and the GLCM textural features (a,b) of the raw dataset and (c,d) the refined dataset, respectively. The larger the mean decrease in accuracy, the more important the feature for the random forest. The prefix of the textural feature names refers to the Sentinel-2 band, and the suffix refers to the respective textural measure. For all abbreviations, the reader is referred to Tables 2 and 3.

2.5. Random Forest Classifier

Supervised machine learning models are inherently linked to the reference data used during training. Model generalisation capacity can be considerably influenced by the size and quality of the training dataset [18,29,98]. Due to our limited reference dataset relative to the large study area, we opted to apply a random forest approach. Random forests, which were introduced by Breiman [95], are an ensemble of decision trees trained on random subsets of input variables and training samples and are known for their robustness across both small and large datasets [10,99–102]. Previous cocoa classification studies with limited data also employed random forests and reported their good performance [43,45,48,67,103].

Another benefit stemming from the application of random forest is the parameterisation stability, which simplifies the optimisation of the hyperparameters [18,102,104]. To maintain the independence of our test dataset, we fine-tuned the hyperparameters based on the validation dataset. The hyperparameter optimisation, the training and the performance assessment of the random forest were performed within the GEE environment [9]. Given that GEE lacks hyperparameter optimisation tools, such as grid search, we applied a heuristic trial-and-error method. We set the number of decision trees to 100 after examining the range of values {100, 200, 300, 500}. The minimum number of samples per leaf was determined to be 10 from the range {1, 5, 10, 15, 20, 30}. The square root of the variables per

node split was selected, which is a commonly used optimisation method and recommended for its effectiveness [98,102].

2.6. Quantitative Assessment and Input Configuration

It is common among classification studies using machine learning models to examine the performance of the derived features and the spectral bands both separately and fused [48,102,105–108]. To identify the most effective combination of predictors, we conducted a comparative analysis of five input configurations:

- The four features selected from the importance ranking, with and without the inclusion of the 12 Sentinel-2 bands;
- The 25 features prior to the selection procedure, with and without the inclusion of the 12 Sentinel-2 bands;
- The 12 Sentinel-2 bands solely.

The performance of the random forest was assessed on each configuration for the raw dataset and the refined dataset using the independent refined test data. The random forest that yielded the highest accuracy was employed for classifying the study area. We calculated the pixel-based measures of overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA) [109] and F1 score based on the confusion matrix (see corresponding Equations (1)–(4)). The OA, which describes the total pixels classified correctly, was deemed suitable for our study because the sizes of the two classification classes were well-balanced. The UA, PA and F1 were expressed per class category. From the cocoa class standpoint, the UA (1 – commission error) represents the proportion of cocoa pixels classified correctly with regard to all pixels classified as cocoa. The PA (1 – omission error) represents the proportion of cocoa pixels classified correctly with regard to the total actual cocoa pixels. The F1, which is the harmonic mean of UA and PA, enables a comparison of models in cases where one model achieves higher UA while the other achieves higher PA, or vice versa. However, we did not compute the kappa coefficient of agreement since its utility and interpretation in classification studies have been questioned [110,111].

$$\text{Overall Accuracy (OA)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (1)$$

$$\text{User's Accuracy (UA)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Producer's Accuracy (PA)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F1} = 2 \times \frac{\text{UA} \times \text{PA}}{\text{UA} + \text{PA}} \quad (4)$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives.

After training the random forest on the training dataset, we assessed the performance ten times by resampling the test data with replacement, i.e., bootstrapping. This approach was adopted to account for the variability associated with the accuracy estimation and to ensure that the accuracy assessment did not rely solely on a specific set of the test data. However, the common practice in similar studies typically involves resampling by repeatedly splitting both training and test datasets to provide estimates of uncertainty [102,107,112,113]. This resampling method demonstrated greater reliability compared with a single sampling iteration [114]. In our study, the division between the training and test datasets was kept fixed to retain the split concept applied for the cocoa and non-cocoa polygons, as described in Section 2.2.2. The resampling for the test data was performed at the polygon level. The mean value and the intervals of one standard deviation were derived for each accuracy measure.

2.7. Qualitative Assessment and Comparison with Previous Cocoa Classification Studies

The last yet integral part of our analysis involved assessing the quality of the cocoa map generated by the best-performing random forest. For this purpose, we conducted visual inspections, which enabled us to critically reflect on the model's generalisation capacity. Through this process, we could identify cases of successful cocoa classification and misclassification, which were findings that quantitative accuracy measures alone could not reveal.

However, the extensive size of our study area (17.4 Mha) made it impractical to conduct a thorough visual inspection across the entire classified area. Documenting every possible visual example in our study was also unfeasible. To address these challenges, we opted to report on a series of illustrative cases of successful and unsuccessful cocoa classification in Côte d'Ivoire and Ghana. Example sites were selected in areas with no training polygons in the proximity, i.e., at least 20 km away. Another requirement for selecting these illustrative examples was that full-sun cocoa trees and other land-use/cover types were visually distinguishable. To this end, we overlaid our classified map on top of very high-spatial-resolution Google Earth images (i.e., usually at submeter resolution), which served as a reliable reference for the actual land type. Only Google Earth images captured in 2021 or in January 2022 at the latest were used to align with our study time frame. We employed the Google Earth Pro desktop version for this purpose [115]. QGIS software version 3.14.1 was utilised for other mapping tasks [116].

To provide a more comprehensive understanding of the potentialities and limitations of our cocoa classification, we visually compared our findings with those of two previous cocoa classification studies carried out by Abu et al. [45] and Kalischek et al. [34]. Both Abu et al. and Kalischek et al. performed binary pixel-based cocoa classification in Côte d'Ivoire and Ghana, though neither of them specified the type of cocoa cultivation they classified, i.e., full-sun cocoa and/or cocoa agroforestry [34,45]. Abu et al. focused on an annual classification for 2019 using a random forest and fused Sentinel-1 and Sentinel-2 data [45]. Kalischek et al. conducted a three-year classification from 2019 to 2021 by employing a convolutional neural network (CNN) and Sentinel-2 imagery [34]. We selected these studies due to their relevance to our study area extent, the similar time frame, the use of optical data, the identical number of classes and the matching spatial resolution. The open access to their classification maps also facilitated our comparative analysis.

3. Results

3.1. Quantitative Assessment

Table 4 displays the accuracy measures scored per dataset and per input configuration. The random forest trained on the four most important VIs and textural features combined with the Sentinel-2 bands achieved the highest OA (mean = 85.1%, standard deviation = $\pm 2.0\%$), PA ($81.9 \pm 3.9\%$) and F1 ($84.6 \pm 2.4\%$) using the refined dataset. The highest PA suggested that more actual cocoa data were correctly classified as cocoa compared with the other models. However, the PA achieved a lower performance compared to UA ($87.5 \pm 0.8\%$), suggesting that the random forest missed classifying more true cocoa pixels. Fusing the 25 features with the Sentinel-2 bands resulted in the highest UA ($88.9 \pm 0.9\%$) and the second-highest OA ($84.7 \pm 1.9\%$) and F1 ($83.8 \pm 2.3\%$).

After removing the four most important features and using the 12 Sentinel-2 bands alone, the mean OA and F1 dropped by approximately 1% and the PA dropped by 2%, while the mean UA increased by 0.3%. Similarly, removing the 25 features from the Sentinel-2 bands resulted in a small drop in accuracy for OA, PA and F1. The handcrafted features therefore improved the classification accuracy when they were combined with the reflectance bands, especially when adding the most important features. Yet, their added value was not big. For the raw dataset, fusion worked in the opposite way; combining the four most important features with the Sentinel-2 bands caused a small accuracy reduction. Nevertheless, the sole use of handcrafted features resulted in the lowest accuracy scores

for both the refined and raw datasets, suggesting that feature engineering alone was not adequate for training the random forest.

Table 4. The accuracy measures evaluated on different input configurations for the refined dataset and the raw dataset. The overall accuracy (OA), user’s accuracy (UA), producer’s accuracy (PA) and F1 were calculated from ten bootstrapping iterations of the refined test data. The mean accuracy \pm one standard deviation intervals are reported. Bold values indicate the highest accuracy per measure. For the UA, PA and F1 of the non-cocoa class, the reader is referred to Table A2 in Appendix A.

Reference Dataset	Input Data	OA (%)	Cocoa		
			UA (%)	PA (%)	F1 (%)
Refined	4 features ¹	80.6 \pm 1.6	82.4 \pm 1.2	77.9 \pm 2.8	80.1 \pm 1.9
	4 features ¹ + 12 Sentinel-2 bands	85.1 \pm 2.0	87.5 \pm 0.8	81.9 \pm 3.9	84.6 \pm 2.4
	25 features ²	82.3 \pm 1.8	87.0 \pm 0.8	75.9 \pm 3.6	81.1 \pm 2.3
	25 features ² + 12 Sentinel-2 bands	84.7 \pm 1.9	88.9 \pm 0.9	79.3 \pm 3.5	83.8 \pm 2.3
	12 Sentinel-2 bands	84.3 \pm 2.4	87.8 \pm 1.1	79.9 \pm 4.6	83.6 \pm 2.9
Raw	4 features ³	82.1 \pm 1.8	80.4 \pm 1.4	84.7 \pm 3.4	82.4 \pm 2.2
	4 features ³ + 12 Sentinel-2 bands	83.4 \pm 2.2	85.0 \pm 1.1	81.2 \pm 4.1	83.0 \pm 2.5
	25 features ⁴	80.7 \pm 3.2	85.5 \pm 1.4	73.9 \pm 6.2	79.2 \pm 4.1
	25 features ⁴ + 12 Sentinel-2 bands	83.9 \pm 2.6	87.7 \pm 1.2	78.8 \pm 5.0	83.0 \pm 3.2
	12 Sentinel-2 bands	83.7 \pm 2.3	85.8 \pm 1.2	80.9 \pm 4.2	83.2 \pm 2.7

¹ NDVI + TCI + B12_savg + B8_savg. ² Eleven VIs + 14 textural features selected from hierarchical clustering (see Figure 2c,d). ³ TCI + NDVI + GNDVI + B6_savg. ⁴ Eleven VIs + 14 textural features selected from hierarchical clustering (see Figure 2a,b).

Employing the same number of inputs for the respective configuration sets of the refined and raw datasets allowed for a wall-to-wall comparison. Also, using the identical test dataset for evaluating the random forest models enabled us to assess the accuracy of the refined and raw datasets under the same conditions. The cocoa quality refinement resulted in an accuracy gain of 1.7% for OA, 2.5% for UA, 0.7% for PA and 1.6% for F1 on average when fusing the four most important features with the Sentinel-2 bands. The refinement approach resulted in an accuracy gain for all input configurations, except when using solely the four most important features. The accuracy gain was greater in some configurations, e.g., the mean F1 increased by 1.9% when using 25 features, while the gain was smaller in other cases, e.g., the mean F1 improved by 0.4% when using solely the Sentinel-2 bands.

The accuracy gain induced by the data refinement confirmed its beneficial impact on model performance. The two-times-bigger raw dataset did not enable the random forest to perform better, which could be attributed to the lower data quality. Most of the accuracy measures obtained from the raw dataset also exhibited wider one standard deviation intervals compared with the refined dataset, indicating greater uncertainty. The corresponding UA, PA and F1 measures for the non-cocoa class are presented in Table A2 in Appendix A to provide a complete overview of both classes.

The random forest trained on the refined dataset using the four most important features and the 12 Sentinel-2 bands was employed to classify the study area. The classified map is presented in Figure 3 depicting full-sun-cocoa-growing areas for 2021 at a 10 m spatial resolution.

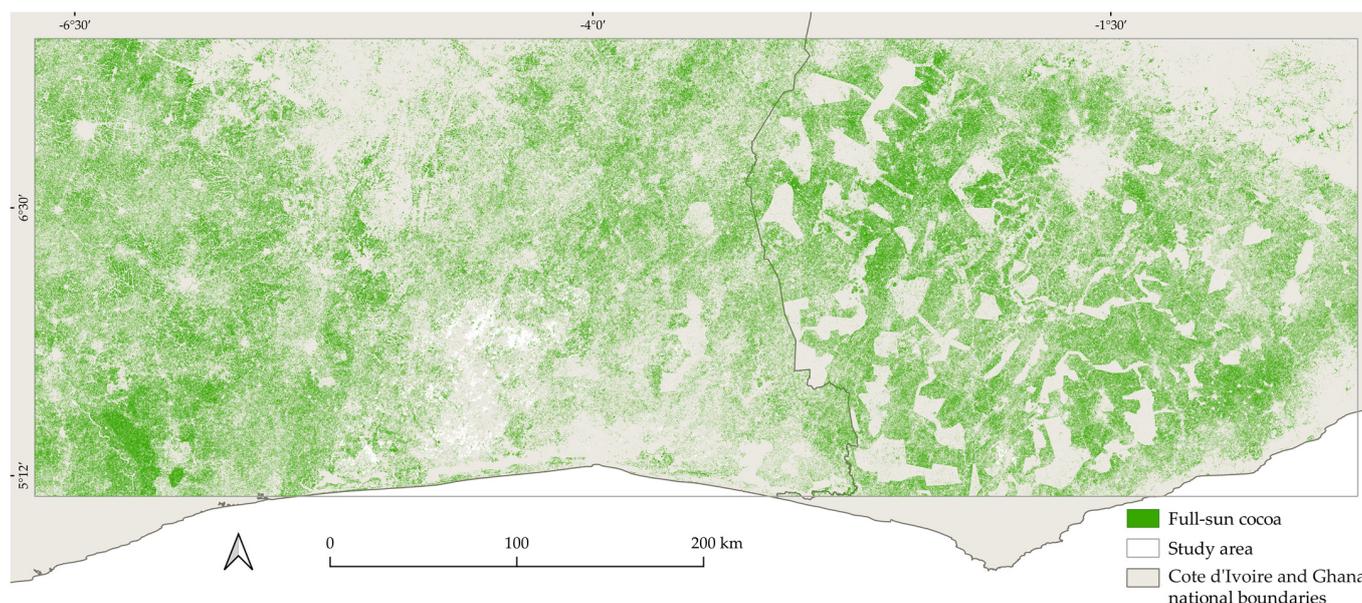


Figure 3. The 2021 classified map of full-sun-cocoa-growing areas at a 10 m spatial resolution over the study area in Côte d'Ivoire and Ghana. White pixels correspond to masked-out pixels. The map is available via <https://my-cocoa-project-102021.projects.earthengine.app/view/cocoa-mapping-west-africa> (accessed on 2 February 2024).

3.2. Qualitative Assessment with Visual Inspection

Figures 4 and 5 illustrate the sites we selected to perform qualitative assessment through visual inspection. In Figure 4, there are examples of successful cocoa classification performed in our study. Figure 5 showcases cocoa misclassification examples, i.e., commission and omission errors. In both figures, each column corresponds to a single zoomed-in site, where the panels in the first row illustrate the 2021 Sentinel-2 median composite image at a 10 m spatial resolution. The panels in the second row include the classified full-sun cocoa by our study. In the third and fourth row, the cocoa classification performed by Kalischek et al. [34] and Abu et al. [45] are illustrated, respectively. Since these authors did not specify the cocoa cultivation type they studied, our legend refers to cocoa in general. All classifications were overlaid on top of very high-spatial-resolution Google Earth images for more reliable and detailed interpretation. The capture dates of the Google Earth images are provided on the respective panels.

In Figure 4a, a primary forest area in Ghana is illustrated with some cocoa encroachment evident therein. Our random forest successfully detected the cocoa trees and delineated them from the forest trees. The CNN model by Kalischek et al. [34] performed in the same manner. In contrast, the random forest by Abu et al. [45] classified in a seemingly arbitrary and random fashion.

Figure 4b exemplifies the typical heterogeneous and complex landscape in Ghana. In Figure 4b, patches of self-standing cocoa trees are mixed with other shade trees, smallholder oil palm trees, bare ground, grassland and a settlement. Our random forest correctly classified the full-sun cocoa, but some sparse shade trees and grassland in between were also classified as cocoa. The remaining land-use/cover was correctly classified as non-cocoa with minor discrepancies. Kalischek et al. [34] classified this site in a similar way with small deviations by assigning shade trees and some bare ground to the cocoa class. However, Abu et al. [45] omitted identifying full-sun cocoa and their model seems to have classified at random.

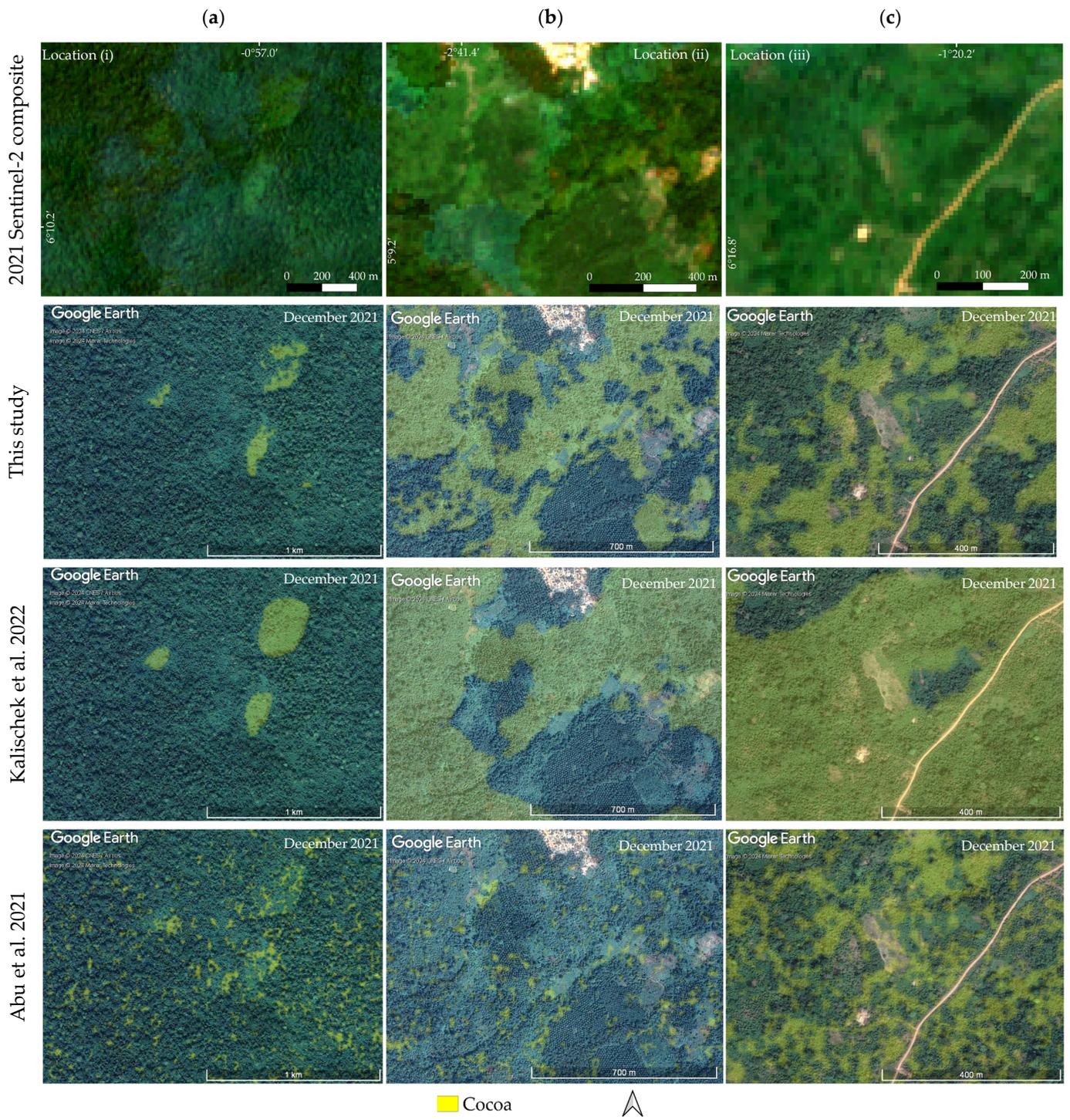


Figure 4. Cont.

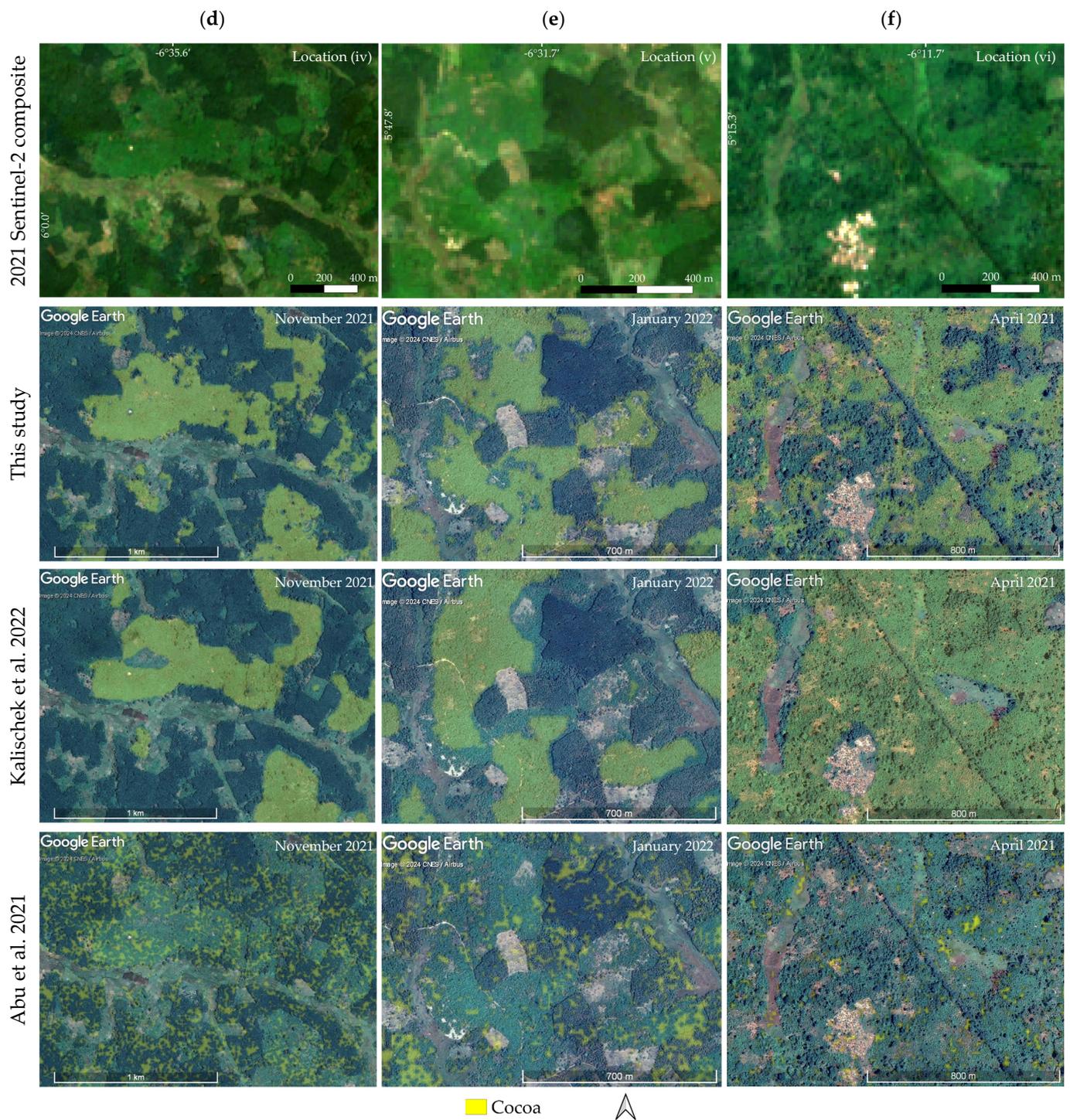


Figure 4. Illustrative example sites in (a–c) Ghana and (d–f) Côte d’Ivoire, where our random forest performed well in classifying full-sun cocoa (second row). Cocoa maps classified by Kalischek et al. [34] and Abu et al. [45] were inspected for comparative analysis (third and fourth row, respectively). The 2021 Sentinel-2 median composite image is provided in true colour (first row). Classified maps were overlaid on top of Google Earth images for finer interpretation (capture dates are attached). Latin numbers refer to the geographic location of the sites relative to the study area, as presented in Figure A2. Legend and north arrow apply to all maps.

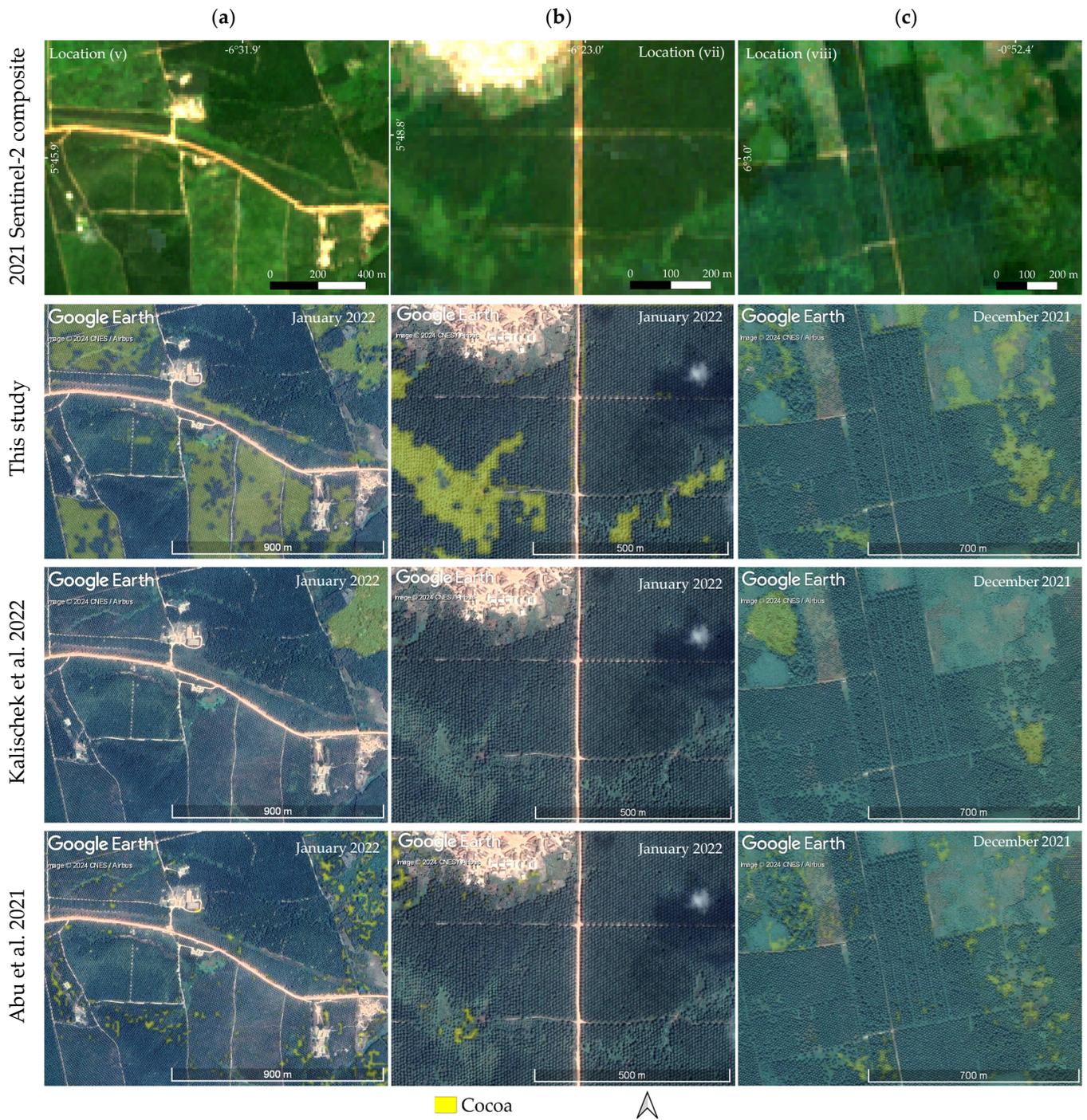


Figure 5. Cont.

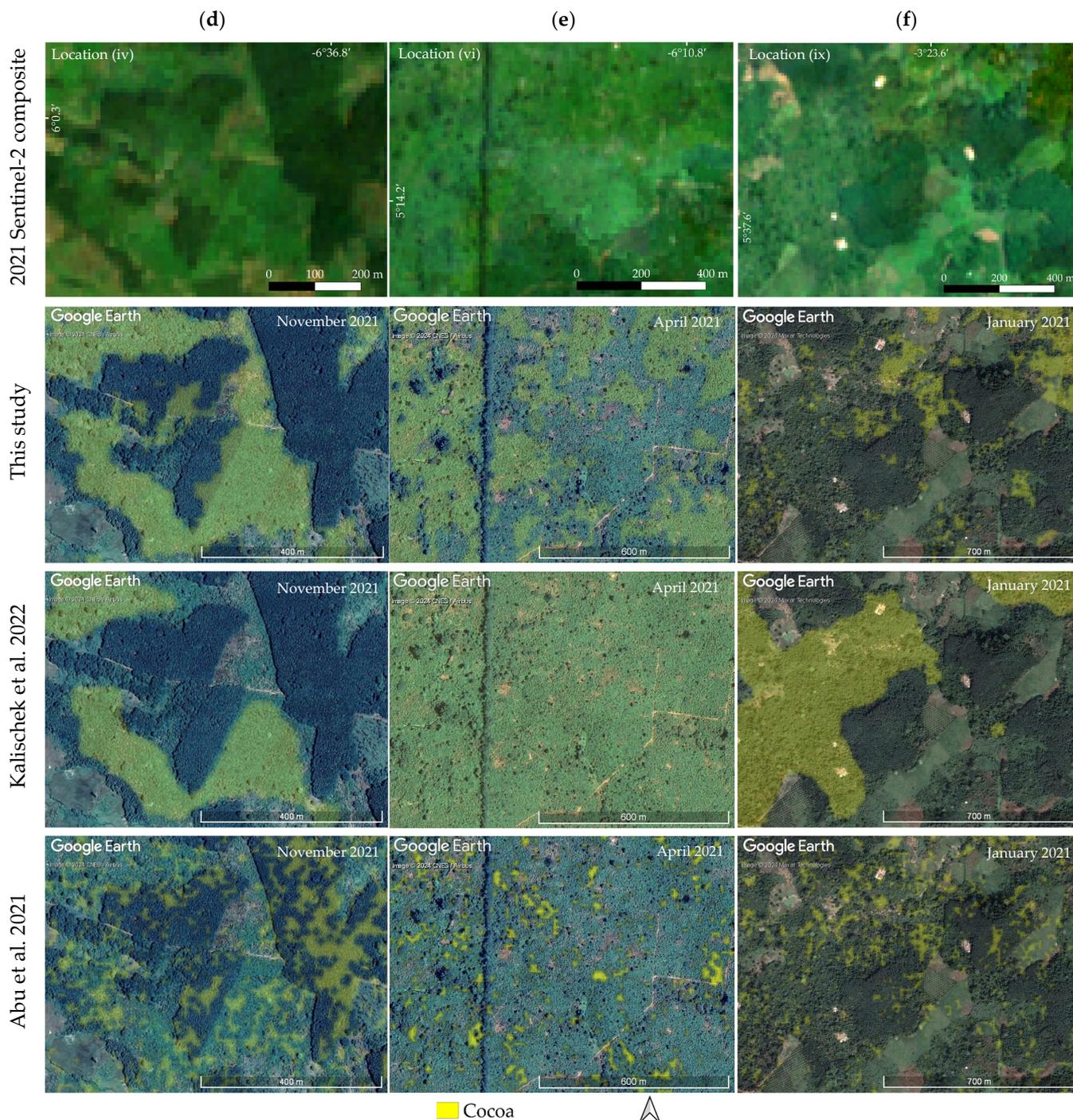


Figure 5. Illustrative example sites in (a,b,d–f) Côte d’Ivoire and (c) Ghana where our random forest performed omission and commission errors in classifying full-sun cocoa (second row). Cocoa maps classified by Kalishek et al. [34] and Abu et al. [45] were inspected for comparative analysis (third and fourth row, respectively). The 2021 Sentinel-2 median composite image is presented in true colour (first row). Classified maps were overlaid on top of Google Earth images for finer interpretation (capture dates are attached). Latin numbers refer to the geographic location of the sites relative to the study area presented in Figure A2. Legend and north arrow apply to all maps.

Figure 4c depicts another site in Ghana with full-sun cocoa mixed with shade trees, forming a dense canopy at certain points (i.e., having a darker green colour in the Sentinel-2 image). This landscape likely constitutes part of a cocoa agroforestry system. Our model successfully segregated the cocoa trees from the closed-canopy shade trees. Also, the bare

ground and the unpaved road were successfully classified as non-cocoa. However, the CNN model by Kalischek et al. [34] did not separate the bare ground and road from cocoa; instead, everything was classified in the cocoa class. The classification by Abu et al. [45] showed inconsistency.

In the sites presented in Figure 4d,e located in Côte d'Ivoire, we identified distinct full-sun cocoa plantations, rubber plantations and bare ground. It is apparent that the West African landscape consists of land-use patches of varying shapes and sizes. In both Figure 4d,e, our random forest delineated the areas of full-sun cocoa accurately, while rubber trees were classified as non-cocoa. A similar classification with small deviations was performed with the CNN model by Kalischek et al. [34]. The random forest applied by Abu et al. [45] seems to have omitted classifying cocoa correctly.

Figure 4f displays another site in Côte d'Ivoire with extensive full-sun cocoa areas mixed with sparse shade trees. A line of remnant forest trees is also visible, along with a nearby settlement. Our random forest successfully classified most of the full-sun cocoa trees located in this site, and most of the forest trees were classified as non-cocoa. Some misclassification occurred for grassland. Kalischek et al. [34] classified both cocoa and shade trees in the cocoa class; apparently their model was capable of detecting cocoa agroforestry. The classification by Abu et al. [45] seems to have generated a random output.

Figure 5a,b show two different Ivorian sites consisting of oil palm fields. In Figure 5a, there are notable commission errors performed by our random forest that confused young oil palm with cocoa, whereas mature oil palm fields were classified correctly as non-cocoa. We confirmed the presence of young oil palm trees from their different reflectance responses in the Sentinel-2 image (i.e., having a lighter green colour than mature oil palm). Additionally, our random forest confused grassland with cocoa. Conversely, the CNN model by Kalischek et al. [34] had a more robust performance and generalised well.

In Figure 5b, a typical industrial oil palm area with a predominantly closed canopy is presented. The industrial plots had a more structured shape where oil palm trees were usually planted in regular rows. Areas with scarce open-canopy oil palm trees were mistakenly detected as cocoa by our random forest. In addition to this, our model misclassified pixels along the oil palm field boundaries in the cocoa class. This issue was more noticeable when examining the same location from a broader perspective in Figure A4a. We observed similar commission errors occurring in other landscapes, where boundary non-cocoa pixels were misclassified as cocoa. These errors were more visually distinct in vegetated areas with straight-line boundaries, e.g., in rubber plantations crossed by paths (see Figure A4b) and in forest areas with logging roads (see Figure A4c). In contrast, the CNN model by Kalischek et al. [34] showed robustness against such commission errors (Figure 5b).

Figure 5c illustrates another site in Ghana with oil palm plantations mixed with grassland. Our random forest successfully detected the closed-canopy oil palm but some sparsely vegetated oil palm trees and grassland were misclassified as cocoa. The models by Kalischek et al. [34] and Abu et al. [45] also confused open-canopy oil palm with cocoa but to a smaller extent.

The site in Figure 5d was in proximity to the area presented in Figure 4d, where a mosaic of rubber and full-sun cocoa trees was located. In this case, our random forest confused some rubber trees with cocoa, but the full-sun cocoa trees were correctly detected. The classification outcome by Kalischek et al. [34] separated rubber trees successfully; however, their model did not classify all cocoa in the area. The classification outcome by Abu et al. [45] seems to be incorrect.

Figure 5e refers to the same location as the one in Figure 4f, where extensive full-sun-cocoa-growing areas are illustrated. Our random forest performed omission errors by misclassifying some full-sun cocoa areas. The shade trees, however, were classified correctly as non-cocoa. Kalischek et al. [34] classified the complete area in the cocoa class. Omission errors were also observed in the study by Abu et al. [45].

Figure 5f illustrates another site with full-sun cocoa mixed with shade trees in Côte d'Ivoire, possibly comprising a cocoa agroforestry area. Most of the full-sun cocoa trees

were omitted by our random forest, while other land types in the surroundings were confused with cocoa. The CNN model by Kalischek et al. [34] correctly classified the cocoa agroforestry, but the scattered farmhouses were also classified as cocoa. The classification by Abu et al. [45] seems to have generated a random output.

4. Discussion

4.1. Cocoa Data Refinement

The quality refinement of the cocoa data had a beneficial impact on the classification accuracy. The mean OA increased by 1.7%; the mean F1 by 1.6%; and the mean UA and PA by 2.5% and 0.7%, respectively, when using the best-performing random forest. An accuracy gain was achieved in almost all input configurations we examined, yet this gain was varied (Table 4). Narrower one standard deviation intervals were achieved for the majority of the scores after the refinement, indicating less uncertainty associated with the accuracy. Even though the refinement resulted in the removal of almost half of the initial cocoa data (Table 1), the size reduction had no negative impact on the performance of the random forest. Instead, the enhancement of the cocoa data quality contributed to the improvement of the model's discrimination capacity.

Other classification studies at larger geographical scales performed more sophisticated data refinement frameworks, which included hierarchically organised datasets based on the reliability level [117], or photointerpretation was undertaken by multiple analysts for increased confidence [22]. In our study, we implemented a simple approach of refining the initial cocoa dataset of uncertain quality by retaining the purest pixels possible that included full-sun cocoa trees. The refinement approach involved manually digitising on Google Earth images of very high-spatial-resolution since the 10 m resolution of the Sentinel-2 image was not sufficient. Our manual labelling was carried out meticulously, yet subjective and biased interpretation could be expected to some extent since the refinement was performed by one photo interpreter [26]. However, our cocoa refinement method was used as a complementary tool to the field measurements and not as a surrogate.

The purpose of the quality refinement was to ensure thematic homogeneity in the cocoa dataset and to mitigate intra-class variability. Nonetheless, it is important to acknowledge that this objective might not be entirely feasible in practice. Elements belonging to the same land-use/cover class may not have homogeneous spectral responses due to, e.g., the health state, the phenology, or the spatial arrangement within the pixel area. This issue can be more challenging in heterogeneous and complex landscapes [118]. When considering solely the spectral behaviour, a pixel's response may resemble more than one land-use/cover class, and, similarly, different classes may have spectral characteristics in common [119]. The Earth's surface is a continuum of various land-use and land-cover classes. However, in hard classification tasks, only one class is allocated to each pixel, as the underlying assumption is that image pixels are pure and represent a discrete land-use/cover class [28].

4.2. Data Sampling in the Context of Limited Reference Data in Large-Scale Study Area

Our study applied a sampling scheme to account for capturing different local particularities for the cocoa and non-cocoa classes. Diverse environmental conditions causing intra-class variability can be expected across large-scale areas. Although cocoa is a perennial crop, it may experience different spectral behaviour based on the health status and the local environmental conditions. For instance, stressed cocoa trees or ones suffering from drought respond spectrally differently from healthy cocoa trees [34,120]. To achieve representative sampling, we allocated polygons from diverse and geographically scattered locations to the training, validation and test datasets. Different splitting designs were applied to each class due to their disparate distribution, i.e., cocoa polygons were unevenly distributed over the study area (Figure 1a), while non-cocoa polygons were more evenly distributed (Figure A1). We divided the non-cocoa polygons based on a gridded block scheme to ensure that the training, validation and test data sampling was performed in disjoint blocks [60]. On the

other hand, we allocated the cocoa polygons into training, validation and test datasets based on spatial homogeneous groups with a minimum distance of 50 km from each other.

Our sampling scheme's objective was also to ensure the spatial independence between the training, validation and test datasets, which is critical for unbiased model performance [59,60,121]. Strategic sampling rather than random sampling is suggested to address issues of spatial autocorrelation, according to which spatially adjacent observations have inherently similar characteristics [60,121]. For our study, random sampling was deemed inappropriate because not only does it ignore spatial autocorrelation but it also requires large-sized datasets to sufficiently represent the classification classes [28]. Previous national-scale studies with millions of reference pixels available trained different random forest models at regional partitions of the study area to capture site-specific particularities and local climatic conditions [113,117]. However, this promising technique is not feasible with limited and scattered reference data.

Given that we did not have randomly sampled and stratified reference data, it was not feasible to calculate an unbiased cocoa area estimation [23], and therefore, we did not report such figures in our study. Pixel counting was not performed, as it would have resulted in an unreliable area estimation.

4.3. Feature Engineering

A series of handcrafted VIs and GLCM textural features were generated, which were examined for their importance with regard to cocoa and non-cocoa classification. The sole use of the four most important VIs and textural features was the weakest input configuration, yielding the lowest accuracy scores (Table 4). When adding the 12 Sentinel-2 bands to these features, both OA and F1 were increased by 4.5% on average using the refined dataset. However, this fusion achieved an average increase of 0.8% for OA and 1% for F1 compared with the sole use of Sentinel-2 bands. Accordingly, the fusion of the Sentinel-2 bands with the 25 initial features (i.e., before selecting the most important ones) improved the mean OA by 0.4% and the F1 by 0.2% compared with the sole use of Sentinel-2 bands. As such, exploring the most important features had an added value in the classification but only when these features were combined with the Sentinel-2 bands. Yet, the accuracy gain from the inclusion of the features to the Sentinel-2 bands relative to the sole use of the Sentinel-2 bands was rather small. This finding is in line with previous classification studies that used a random forest [102,107].

Textural features were shown to be particularly beneficial for classification studies in heterogeneous areas [69,105], and especially in the complex and fragmented landscape of West Africa, where smallholder farms of irregular shapes predominate [48,108]. However, the window size plays a key role in the effectiveness of texture, and its selection mainly depends on the image resolution and the homogeneity or heterogeneity of the land-cover class [122]. The window size cannot be too small, only partially capturing the land-cover class, or too big, covering the neighbouring class [123]. In our study, we applied a 3×3 -pixel window size following the practice by Abu et al. [45], but this size might have been too small considering the average cocoa plot area of 2.3 hectares and the 10 m spatial resolution of the Sentinel-2 image. Previous studies found that a window size of 5×5 -pixel was more suitable for detecting cocoa agroforestry and coffee using Sentinel-1 images [35,48]. Therefore, the effectiveness of larger window sizes and/or the combination of various sizes for generating GLCM textural features could be investigated in future cocoa studies using optical data.

4.4. Image Compositing

The number of cloud-free images available at a given location can influence the map quality [113]. Creating a composite image from optical data is considered a suitable solution when the study area is large and susceptible to persistent cloud coverage [124]. Our study managed to generate a quality and cloud-free Sentinel-2 median composite to capture the cocoa temporal dynamics in 2021. However, it might be expected that the dry and wet seasonality did not equally contribute to the composite image, and as a result, the annual

cocoa dynamics might not have been represented appropriately. Future cocoa studies in West Africa could further explore the effectiveness of using multiple composites from the dry and wet seasons and/or deriving features from time-series images to account for the seasonal vegetation and the classes prone to changes. Another shortcoming of our compositing method is that the median reflectance values of neighbouring pixels might have been retrieved from different image scenes. This could have resulted in deviating values for pixels that were supposed to be classified in the same class but were assigned to different ones.

In total, three Sentinel-2 orbits were required to mosaic our composite image covering the study area extent. However, satellite orbits have different revisit times, and consequently, the swath areas from adjacent orbits are captured under different viewing conditions. This issue, which was also encountered in other studies [62,113], caused streak artefacts in the Sentinel-2 composite image (Figure A2) and the final cocoa map (Figure 3); however, these streaks are not evident at a local zoom-in level.

4.5. Qualitative Assessment with Visual Inspection

Commonly, a classification map is evaluated based on the accuracy reported [28]. In our study, we conducted a qualitative assessment by visually inspecting our cocoa map as a complement to the accuracy measures. The aim was to gain insight into our model's discriminating ability, which could not be revealed solely from the quantitative accuracy measures, although they were evaluated on an unseen and independent test dataset. The accuracy measures can only assess the degree of agreement with the reference data and not the actual correspondence to reality [28]. Therefore, we visually inspected our classification map overlaid on very high-spatial-resolution images to enable a finer interpretation. For feasibility reasons, we provided some illustrative cases of successful cocoa classification and misclassification. However, we need to bear in mind that such a qualitative approach cannot provide an exhaustive overview of the model's generalisation capacity. Instead, we could only point out some of the potentialities and weaknesses of the classified map, as derived from the individual example sites.

Several land-use/cover studies have incorporated visual examples of their classification outcome compared with the satellite input image or images with very high-spatial-resolution [33,35,107,113,125,126]. The authors of [113] performed an extensive visual inspection approach to communicate anomalous patterns and errors for their national-scale map, and the authors of [126] incorporated many visual examples to comment on the quality of their continental-scale map. The global study by [33] also provided site examples where their model succeeded or failed to detect oil palm plantations. Communicating the limitations of the classified outcome to the users with regard to the intended map purposes is of great importance [28]. However, we noticed that recently published cocoa studies provided only their final thematic map with no focus on reflecting on the quality of the classification outcome [45–48,67]. Only the authors of [34,43] illustrated some classification examples but to a small extent.

With a visual inspection, we showcased that our best-performing random forest was able to generalise well in some sites where full-sun cocoa trees were classified successfully and closed-canopy shade trees were delineated (Figure 4). In these examples, our random forest performed similarly to the CNN model developed by Kalischek et al. [34]. Nonetheless, we detected a series of cocoa commission errors in our classified map (Figure 5a–d). Grassland, young oil palm trees and sparse vegetation constituted, among others, the main confusion with full-sun cocoa trees based on the illustrative examples. This is consistent with previous multi-class studies using optical data in Ghana, where cocoa was confused with oil palm and scrubland/grassland [43,47]. A drawback of our binary classification was the lack of insight into inter-class confusion, whereas multi-class classification could provide a clearer understanding. Another type of commission error was observed across class borders and in the transitional zones between classes at some sites (Figure A4). These types of errors may correspond to mixed pixels. Other studies performed polygon erosion

in their reference dataset to prevent errors caused by mixed pixels [102,113,117]. We did not apply this technique because of the limited area of our cocoa polygons.

The visual inspection also enabled us to compare the discriminating capacity between a conventional machine learning classifier, i.e., random forest, and a deep learning model, i.e., CNN. In this way, we confirmed the presumed superiority of deep learning models that inherently take into account the context and spatial surroundings, as opposed to the traditional classifiers, which consider only the pixel-based information retrieved from handcrafted input data [127]. The major shortcoming of deep learning models is, however, their requirement of a large training dataset for optimal performance [19]. Kalischek et al. utilised more than 100,000 cocoa farms to train a CNN model [34]. In contrast, we had 68 and 11 cocoa polygons available for the training and test dataset, from which we refined 4437 and 723 pixels, respectively. Yet, our random forest that was trained on a limited dataset exhibited relatively good potential, as indicated by the derived accuracy measures and the visual inspection. For future cocoa studies that experience limited and inadequately representative data, an active learning scheme could be explored [128], following the recommendations of previous studies [17,129]. Active learning is a concept in the machine learning field that does not require a large original training dataset, as a more effective dataset is built interactively by the model and the analyst by sampling informative pixels from the unlabeled dataset pool [130].

4.6. Comparison with Other Cocoa Maps and Implications for Reporting on Area Estimation

The comparative analysis of our cocoa map with the maps generated by Kalischek et al. [34] and Abu et al. [45] showed that solely interpreting the accuracy scores can be misleading for the map's quality. For instance, in terms of the PA for the cocoa class, the scores achieved in our study and these two studies were approximately in agreement: our model achieved 81.9% on average, while Kalischek et al. and Abu et al. reported 87.2% and 82.9%, respectively [34,45]. Yet, with a visual inspection, we discovered disparities between the three cocoa maps, highlighting the added value of the qualitative visual assessment. The CNN model by Kalischek et al. [34] demonstrated robust performance in detecting full-sun cocoa and cocoa agroforestry. Conversely, the random forest by Abu et al. [45] showed a random and poor performance despite the adequate accuracy scores they reported. However, the findings by Abu et al. [45] were consulted in reports with regard to the cocoa area estimation [131,132], while their map has been distributed in the web-based platform of Copernicus [133]. It becomes apparent that there is a seeming objectivity in quantitative accuracy measures and their interpretation should be approached with caution [28]. Therefore, the synergistic application of quantitative and qualitative accuracy assessment is recommended for communicating map quality in a comprehensive way. Clarity on the (in)accuracies of the classification map is necessary, especially when the outcome is to be used in high-level reporting against global and national policies and commitments, such as zero deforestation.

5. Conclusions

In our study, we aimed at a good trade-off between the limited reference data and an effective cocoa classification using a random forest. To achieve this, we undertook reference data refinement, a sampling strategy and feature selection. Applying quality refinement to the cocoa data contributed to the accuracy improvement. As a complementary tool to the quantitative accuracy measures, we performed a qualitative assessment by visually inspecting our classification outcome overlaid on very high-resolution images. The visual inspection revealed potentialities and limitations in random forest performance, insights that were not evident from the accuracy measures alone. Therefore, it is critical for studies with limited reference data and with large, complex areas to classify to optimise the pre-processing methodology steps and to critically interpret the classification outcome. This can help to clearly communicate the value of the classified map to the end users to leverage it accordingly.

Author Contributions: Conceptualisation, A.M., J.R. and N.M.; methodology, A.M., J.R. and N.M.; data acquisition, E.R. and M.S.; data curation, N.M.; software, N.M.; writing—original draft preparation, N.M.; writing—review and editing, A.M., E.R., M.S., J.R. and N.M.; visualisation, N.M.; supervision, A.M. and J.R.; funding acquisition, J.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Norway’s Climate and Forest Initiative (NICFI), the US Government’s SilvaCarbon program, and the Netherlands Science Funding PIPP project MINERVA No. KNW19001. Cocoa field data were contributed through the CocoaSoils project, which was funded by the Norwegian Agency for Development Cooperation (NORAD), grant number RF-17/0009-Cocoasoils. M. Sassen also acknowledges support from the UK Research and Innovation’s Global Challenges Research Fund (UKRI GCRF) through the Trade, Development and Environment Hub project (project number ES/S008160/1).

Data Availability Statement: The cocoa data used in this study are subject to a non-disclosure agreement. They are available from the corresponding author upon reasonable request. The generated maps, i.e., binary classification map, probability score map and Sentinel-2 composite image, are available for visualisation and download in the following Google Earth Engine application: <https://my-cocoa-project-102021.projects.earthengine.app/view/cocoa-mapping-west-africa> (accessed on 2 February 2024).

Acknowledgments: The authors would like to thank the three anonymous reviewers for their constructive feedback.

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

Appendix A

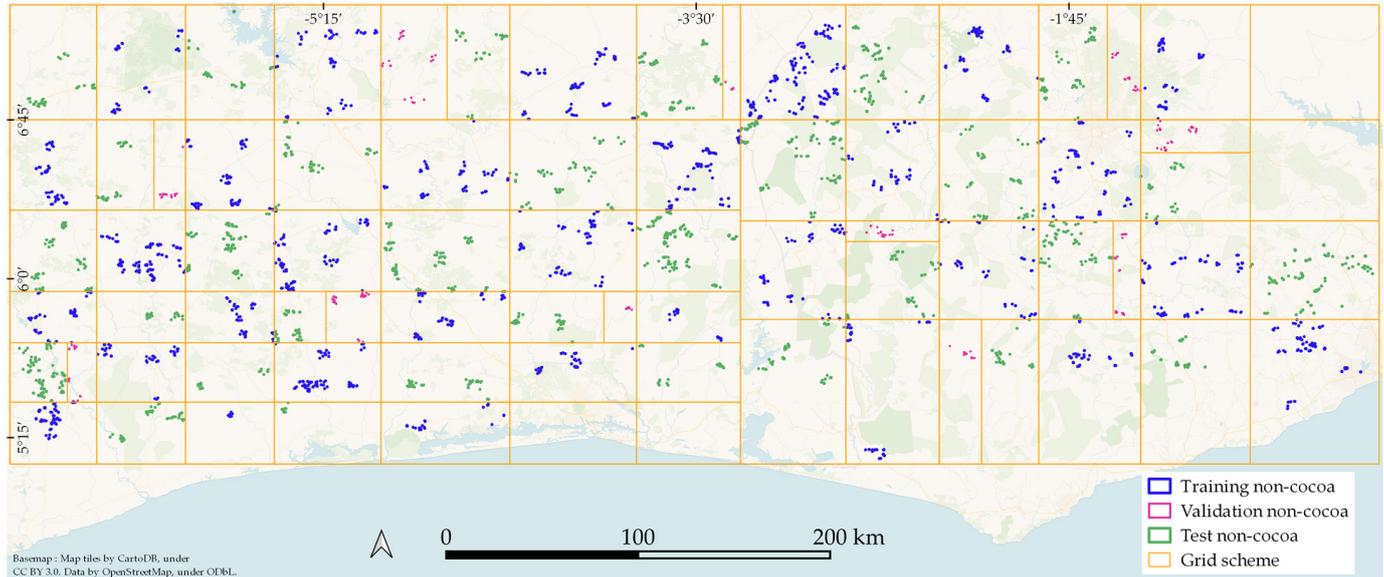


Figure A1. We curated 3311 non-cocoa polygons and split them into training, validation and test datasets based on gridded block sampling scheme to account for spatial autocorrelation and to ensure diverse regional coverage.

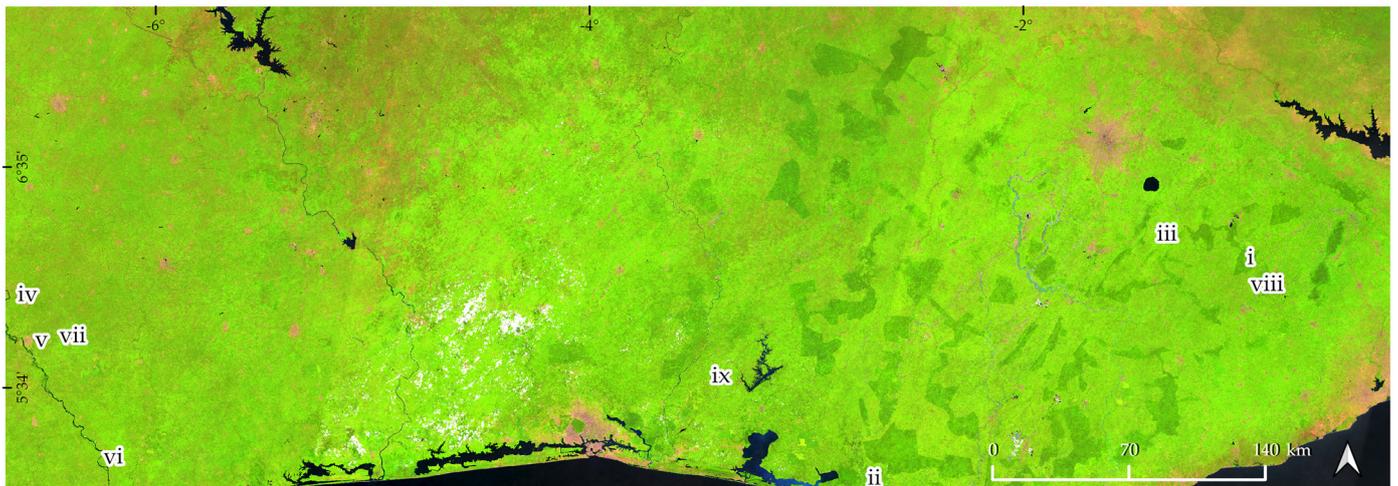


Figure A2. The 2021 Sentinel-2 median composite image in false colour (red: B11, green: B8, blue: B4) over the study area. Latin numbers refer to the geographic locations of the example sites illustrated in Figures 4 and 5. White pixels correspond to masked-out pixels.

Table A1. Parameter values set for cloud and cloud shadow masking with s2cloudless algorithm.

s2cloudless Parameters *	Values
Cloud filter	100
Cloud probability threshold	20
Near-infrared dark threshold	0.2
Cloud projection distance	1.5
Buffer	50

* Parameter names follow the nomenclature defined in s2cloudless algorithm [66].

Table A2. The user's accuracy (UA), producer's accuracy (PA) and F1 for the non-cocoa class derived from ten bootstrapping iterations of the refined test data. The mean accuracy \pm one standard deviation intervals are reported. Bold values indicate the highest accuracy per measure.

Reference Dataset	Input Data	Non-Cocoa		
		UA (%)	PA (%)	F1 (%)
Refined	4 features	79.0 \pm 2.2	83.3 \pm 1.1	81.1 \pm 1.4
	4 features + 12 Sentinel-2 bands	83.0 \pm 2.9	88.3 \pm 0.7	85.6 \pm 1.6
	25 features	78.6 \pm 2.5	88.6 \pm 0.6	83.3 \pm 1.4
	25 features + 12 Sentinel-2 bands	81.3 \pm 2.6	90.1 \pm 0.7	85.5 \pm 1.6
	12 Sentinel-2 bands	81.6 \pm 3.5	88.8 \pm 0.8	85.0 \pm 2.0
Raw	4 features	84.0 \pm 2.9	79.5 \pm 1.1	81.6 \pm 1.5
	4 features + 12 Sentinel-2 bands	82.0 \pm 3.2	85.6 \pm 0.8	83.8 \pm 1.8
	25 features	77.0 \pm 4.2	88.0 \pm 0.7	82.0 \pm 2.5
	25 features + 12 Sentinel-2 bands	81.0 \pm 3.7	89.0 \pm 0.9	85.0 \pm 2.1
	12 Sentinel-2 bands	82.0 \pm 3.3	86.5 \pm 0.9	84.2 \pm 2.0

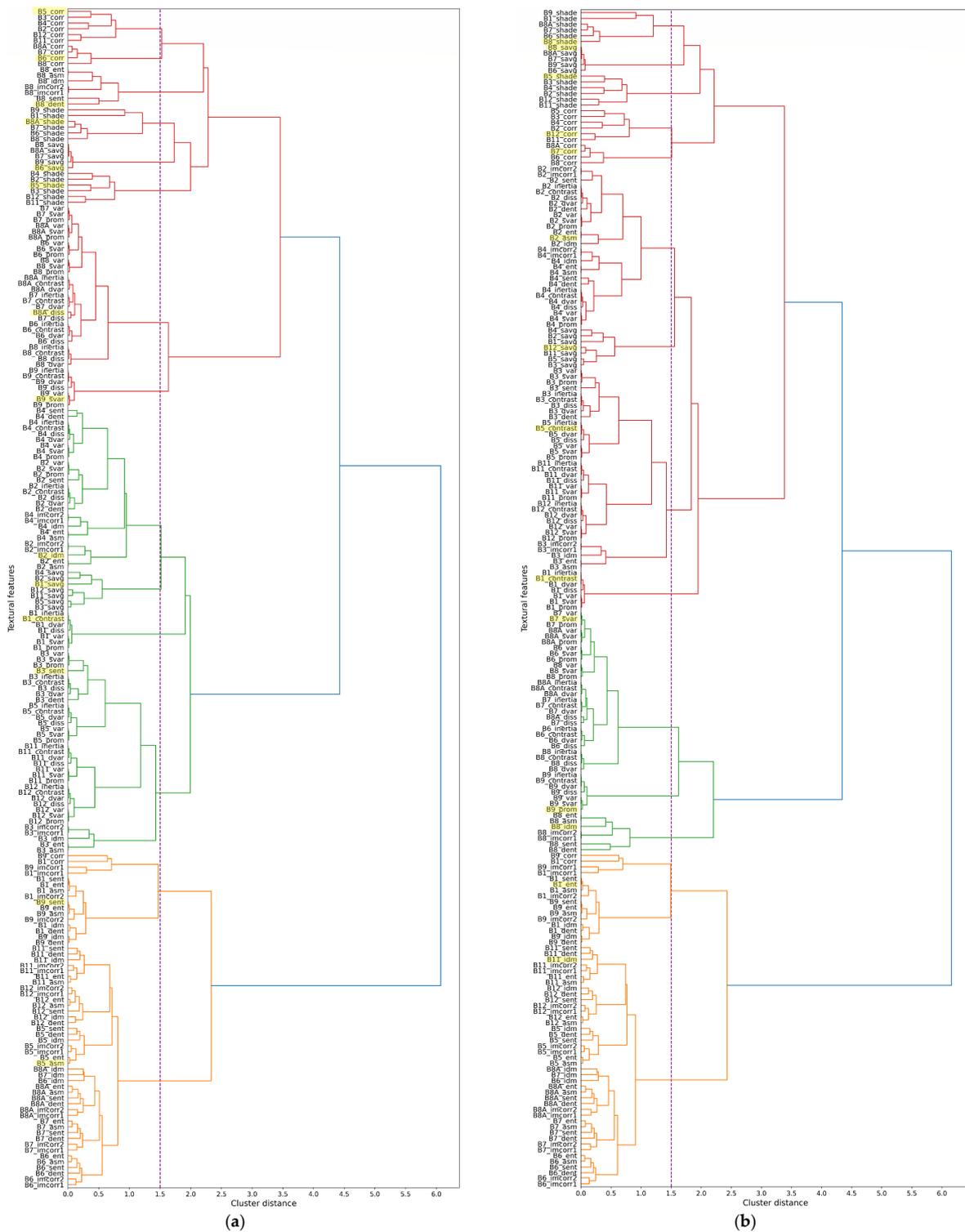


Figure A3. Generated GLCM-based textural features were grouped into clusters after performing hierarchical clustering. In the dendrograms derived from (a) the raw dataset and (b) the refined dataset, the small distances indicate higher correlation between features. We cut both dendrograms at 1.5 (dashed line) and 14 clusters were yielded in each case. From each cluster, one textural feature was randomly selected (highlighted with yellow), resulting in 14 features. These features were then evaluated for their relative importance. The prefix of the feature names refers to the Sentinel-2 band and the suffix refers to the respective textural measure.

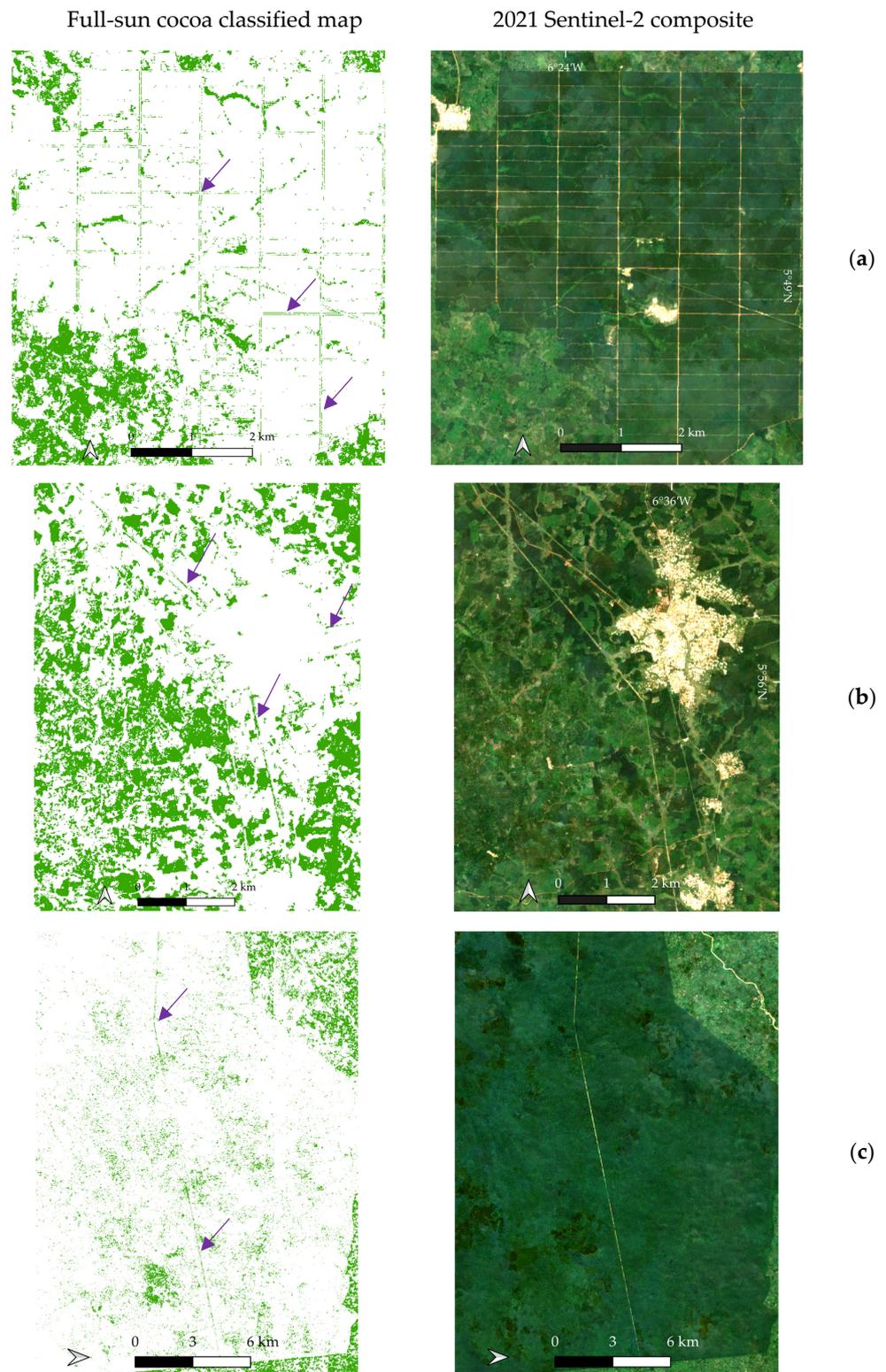


Figure A4. Illustrative examples of cocoa commission errors detected in the visual inspection of our classified map in (a,b) Côte d'Ivoire and (c) Ghana. These errors were mainly found across class boundaries and in the transitional zones between classes (i.e., they are more distinct on straight-line boundaries covered with grassland). Purple arrows roughly indicate the error locations. Green pixels on the map represent the classified full-sun cocoa.

References

- Joshi, N.; Baumann, M.; Ehammer, A.; Fensholt, R.; Grogan, K.; Hostert, P.; Jepsen, M.R.; Kuemmerle, T.; Meyfroidt, P.; Mitchard, E.T.A.; et al. A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. *Remote Sens.* **2016**, *8*, 70. [CrossRef]
- Mayaux, P.; Pekel, J.-F.; Desclée, B.; Donnay, F.; Lupi, A.; Achard, F.; Clerici, M.; Bodart, C.; Brink, A.; Nasi, R.; et al. State and Evolution of the African Rainforests between 1990 and 2010. *Phil. Trans. R. Soc. B* **2013**, *368*, 20120300. [CrossRef] [PubMed]
- Reiche, J.; Mullissa, A.; Slagter, B.; Gou, Y.; Tsendbazar, N.-E.; Odongo-Braun, C.; Vollrath, A.; Weisse, M.J.; Stolle, F.; Pickens, A.; et al. Forest Disturbance Alerts for the Congo Basin Using Sentinel-1. *Environ. Res. Lett.* **2021**, *16*, 024005. [CrossRef]
- Tuanmu, M.-N.; Jetz, W. A Global 1-km Consensus Land-Cover Product for Biodiversity and Ecosystem Modelling. *Glob. Ecol. Biogeogr.* **2014**, *23*, 1031–1045. [CrossRef]
- Szantoi, Z.; Geller, G.N.; Tsendbazar, N.-E.; See, L.; Griffiths, P.; Fritz, S.; Gong, P.; Herold, M.; Mora, B.; Obregón, A. Addressing the Need for Improved Land Cover Map Products for Policy Support. *Environ. Sci. Policy* **2020**, *112*, 28–35. [CrossRef]
- Norway's International Climate and Forest Initiative (NICFI). New Satellite Images to Allow Anyone, Anywhere, to Monitor Tropical Deforestation. Available online: <https://www.nicfi.no/current/new-satellite-images-to-allow-anyone-anywhere-to-monitor-tropical-deforestation/> (accessed on 27 September 2023).
- Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]
- Torres, R.; Snoeij, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.; Floury, N.; Brown, M.; et al. GMES Sentinel-1 Mission. *Remote Sens. Environ.* **2012**, *120*, 9–24. [CrossRef]
- Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
- Gong, P.; Liu, H.; Zhang, M.; Li, C.; Wang, J.; Huang, H.; Clinton, N.; Ji, L.; Li, W.; Bai, Y.; et al. Stable Classification with Limited Sample: Transferring a 30-m Resolution Sample Set Collected in 2015 to Mapping 10-m Resolution Global Land Cover in 2017. *Sci. Bull.* **2019**, *64*, 370–373. [CrossRef] [PubMed]
- Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **2013**, *342*, 850–853. [CrossRef] [PubMed]
- Ge, J.; Qi, J.; Lofgren, B.M.; Moore, N.; Torbick, N.; Olson, J.M. Impacts of Land Use/Cover Classification Accuracy on Regional Climate Simulations. *J. Geophys. Res.* **2007**, *112*, D05107. [CrossRef]
- McMahon, G. Consequences of Land-Cover Misclassification in Models of Impervious Surface. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 1343–1353. [CrossRef]
- Li, C.; Wang, J.; Wang, L.; Hu, L.; Gong, P. Comparison of Classification Algorithms and Training Sample Sizes in Urban Land Classification with Landsat Thematic Mapper Imagery. *Remote Sens.* **2014**, *6*, 964–983. [CrossRef]
- Rogan, J.; Franklin, J.; Stow, D.; Miller, J.; Woodcock, C.; Roberts, D. Mapping Land-Cover Modifications over Large Areas: A Comparison of Machine Learning Algorithms. *Remote Sens. Environ.* **2008**, *112*, 2272–2283. [CrossRef]
- Jin, H.; Stehman, S.V.; Mountrakis, G. Assessing the Impact of Training Sample Selection on Accuracy of an Urban Classification: A Case Study in Denver, Colorado. *Int. J. Remote Sens.* **2014**, *35*, 2067–2081. [CrossRef]
- Heydari, S.S.; Mountrakis, G. Effect of Classifier Selection, Reference Sample Size, Reference Class Distribution and Scene Heterogeneity in Per-Pixel Classification Accuracy Using 26 Landsat Sites. *Remote Sens. Environ.* **2018**, *204*, 648–658. [CrossRef]
- Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [CrossRef]
- Ball, J.E.; Anderson, D.T.; Chan, C.S. Comprehensive Survey of Deep Learning in Remote Sensing: Theories, Tools, and Challenges for the Community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [CrossRef]
- Ramezan, C.A.; Warner, T.A.; Maxwell, A.E.; Price, B.S. Effects of Training Set Size on Supervised Machine-Learning Land-Cover Classification of Large-Area High-Resolution Remotely Sensed Data. *Remote Sens.* **2021**, *13*, 368. [CrossRef]
- Copass, C.; Antonova, N.; Kennedy, R. Comparison of Office and Field Techniques for Validating Landscape Change Classification in Pacific Northwest National Parks. *Remote Sens.* **2019**, *11*, 3. [CrossRef]
- Zhao, Y.; Gong, P.; Yu, L.; Hu, L.; Li, X.; Li, C.; Zhang, H.; Zheng, Y.; Wang, J.; Zhao, Y.; et al. Towards a Common Validation Sample Set for Global Land-Cover Mapping. *Int. J. Remote Sens.* **2014**, *35*, 4795–4814. [CrossRef]
- Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good Practices for Estimating Area and Assessing Accuracy of Land Change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [CrossRef]
- Asare, R.; Markussen, B.; Asare, R.A.; Anim-Kwapong, G.; Ræbild, A. On-Farm Cocoa Yields Increase with Canopy Cover of Shade Trees in Two Agro-Ecological Zones in Ghana. *Clim. Dev.* **2019**, *11*, 435–445. [CrossRef]
- Hainmueller, J.; Hiscox, M.J.; Tampe, M. *Sustainable Development for Cocoa Farmers in Ghana*; MIT and Harvard University: Cambridge, MA, USA, 2011; p. 66.
- Powell, R.L.; Matzke, N.; De Souza, C.; Clark, M.; Numata, I.; Hess, L.L.; Roberts, D.A. Sources of Error in Accuracy Assessment of Thematic Land-Cover Maps in the Brazilian Amazon. *Remote Sens. Environ.* **2004**, *90*, 221–234. [CrossRef]

27. Congalton, R.G.; Green, K. A Practical Look at the Sources of Confusion in Error Matrix Generation. *Photogramm. Eng. Remote Sens.* **1993**, *59*, 641–644.
28. Foody, G.M. Status of Land Cover Classification Accuracy Assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201. [CrossRef]
29. Foody, G.M.; Pal, M.; Rocchini, D.; Garzon-Lopez, C.X.; Bastin, L. The Sensitivity of Mapping Methods to Reference Data Quality: Training Supervised Image Classifications with Imperfect Reference Data. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 199. [CrossRef]
30. Mellor, A.; Boukir, S.; Haywood, A.; Jones, S. Exploring Issues of Training Data Imbalance and Mislabelling on Random Forest Performance for Large Area Land Cover Classification Using the Ensemble Margin. *ISPRS J. Photogramm. Remote Sens.* **2015**, *105*, 155–168. [CrossRef]
31. Foody, G.M. Ground Reference Data Error and the Mis-Estimation of the Area of Land Cover Change as a Function of its Abundance. *Remote Sens. Lett.* **2013**, *4*, 783–792. [CrossRef]
32. Halladin-Dąbrowska, A.; Kania, A.; Kopeć, D. The t-SNE Algorithm as a Tool to Improve the Quality of Reference Data Used in Accurate Mapping of Heterogeneous Non-Forest Vegetation. *Remote Sens.* **2020**, *12*, 39. [CrossRef]
33. Descals, A.; Wich, S.; Meijaard, E.; Gaveau, D.L.A.; Peedell, S.; Szantoi, Z. High-Resolution Global Map of Smallholder and Industrial Closed-Canopy Oil Palm Plantations. *Earth Syst. Sci. Data* **2021**, *13*, 1211–1231. [CrossRef]
34. Kalischek, N.; Lang, N.; Renier, C.; Daudt, R.; Addoah, T.; Thompson, W.; Blaser-Hart, W.; Garrett, R.; Wegner, J. Satellite-Based High-Resolution Maps of Cocoa Planted Area for Côte d’Ivoire and Ghana. *arXiv* **2022**, arXiv:2206.06119. [CrossRef]
35. Maskell, G.; Chemura, A.; Nguyen, H.; Gornott, C.; Mondal, P. Integration of Sentinel Optical and Radar Data for Mapping Smallholder Coffee Production Systems in Vietnam. *Remote Sens. Environ.* **2021**, *266*, 112709. [CrossRef]
36. Somarriba, E.; López-Sampson, A. *Coffee and Cocoa Agroforestry Systems: Pathways to Deforestation, Reforestation, and Tree Cover Change*; The World Bank: Washington, DC, USA, 2018; p. 50.
37. Asare, R.; Afari-Sefa, V.; Osei-Owusu, Y.; Pabi, O. Cocoa Agroforestry for Increasing Forest Connectivity in a Fragmented Landscape in Ghana. *Agroforest. Syst.* **2014**, *88*, 1143–1156. [CrossRef]
38. Barima, Y.; Kouakou, A.; Bamba, I.; Yao Charles, S.; Godron, M.; Andrieu, J.; Bogaert, J. Cocoa Crops are Destroying the Forest Reserves of the Classified Forest of Haut-Sassandra (Ivory Coast). *Glob. Ecol. Conserv.* **2016**, *8*, 85–98. [CrossRef]
39. Cocoa and Forests Initiative. Cocoa and Forests Initiative. Available online: <https://www.idhsustainabletrade.com/initiative/cocoa-and-forests/> (accessed on 27 September 2023).
40. Mondeléz International Cocoa Life. Progress Blog. Available online: <https://www.cocoalife.org/progress> (accessed on 27 September 2023).
41. Nestlé Cocoa Plan. Towards Forest Positive Cocoa Progress Report 2023. Available online: <https://www.nestlecocoaplan.com/article-towards-forest-positive-cocoa-0> (accessed on 27 September 2023).
42. The European Parliament; The Council of the European Union. *Regulation (EU) 2023/1115 of the European Parliament and of the Council of 31 May 2023 on the Making Available on the Union Market and the Export from the Union of Certain Commodities and Products Associated with Deforestation and Forest Degradation and Repealing Regulation (EU) No 995/2010*; European Union: Brussels, Belgium, 2023; pp. 206–247.
43. Ashiagbor, G.; Forkuo, E.K.; Asante, W.A.; Acheampong, E.; Quaye-Ballard, J.A.; Boamah, P.; Mohammed, Y.; Foli, E. Pixel-Based and Object-Oriented Approaches in Segregating Cocoa from Forest in the Juabeso-Bia Landscape of Ghana. *Remote Sens. Appl. Soc. Environ.* **2020**, *19*, 100349. [CrossRef]
44. Sassen, M.; Van Soesbergen, A.; Arnell, A.P.; Scott, E. Patterns of (Future) Environmental Risks from Cocoa Expansion and Intensification in West Africa Call for Context Specific Responses. *Land Use Policy* **2022**, *119*, 106142. [CrossRef]
45. Abu, I.-O.; Szantoi, Z.; Brink, A.; Robuchon, M.; Thiel, M. Detecting Cocoa Plantations in Côte d’Ivoire and Ghana and their Implications on Protected Areas. *Ecol. Indic.* **2021**, *129*, 107863. [CrossRef] [PubMed]
46. Asubonteng, K.; Pfeffer, K.; Ros-Tonen, M.; Verbesselt, J.; Baud, I. Effects of Tree-Crop Farming on Land-Cover Transitions in a Mosaic Landscape in the Eastern Region of Ghana. *Environ. Manag.* **2018**, *62*, 529–547. [CrossRef]
47. Benefoh, D.T.; Villamor, G.B.; Van Noordwijk, M.; Borgemeister, C.; Asante, W.A.; Asubonteng, K.O. Assessing Land-Use Typologies and Change Intensities in a Structurally Complex Ghanaian Cocoa Landscape. *Appl. Geogr.* **2018**, *99*, 109–119. [CrossRef]
48. Numbisi, F.N.; Van Coillie, F.M.B.; De Wulf, R. Delineation of Cocoa Agroforests Using Multiseason Sentinel-1 SAR Images: A Low Grey Level Range Reduces Uncertainties in GLCM Texture-Based Mapping. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 179. [CrossRef]
49. Läderach, P.; Martinez Valle, A.; Schroth, G.; Castro, N. Predicting the Future Climatic Suitability for Cocoa Farming of the World’s Leading Producer Countries, Ghana and Côte d’Ivoire. *Clim. Chang.* **2013**, *119*, 841–854. [CrossRef]
50. Schroth, G.; Läderach, P.; Martinez Valle, A.; Bunn, C. From Site-Level to Regional Adaptation Planning for Tropical Commodities: Cocoa in West Africa. *Mitig. Adapt. Strateg. Glob. Chang.* **2017**, *22*, 903–927. [CrossRef]
51. FAO; ICRISAT; CIAT. *Climate-Smart Agriculture in Côte d’Ivoire. CSA Country Profiles for Africa Series*; FAO: Rome, Italy, 2018; p. 23.
52. Ministry of Food and Agriculture (MOFA); Statistics, Research and Information Directorate (SRID). *Agriculture in Ghana: Facts and Figures 2019*; MOFA: Accra, Ghana, 2020.
53. Forestry Commission. *Ghana REDD+ Strategy 2016–2035*; Forestry Commission: Accra, Ghana, 2016; p. 101.
54. FAO. *State of the World’s Forests 2016. Forests and Agriculture: Land-Use Challenges and Opportunities*; FAO: Rome, Italy, 2016.
55. Ruf, F.O. The Myth of Complex Cocoa Agroforests: The Case of Ghana. *Hum. Ecol.* **2011**, *39*, 373–388. [CrossRef]

56. Duguma, B.; Gockowski, J.; Bakala, J. Smallholder Cacao (*Theobroma Cacao* Linn.) Cultivation in Agroforestry Systems of West and Central Africa: Challenges and Opportunities. *Agrofor. Syst.* **2001**, *51*, 177–188. [[CrossRef](#)]
57. Laven, A.; Bymolt, R.; Tyszler, M. *Demystifying the Cocoa Sector in Ghana and Côte d'Ivoire*; The Royal Tropical Institute (KIT): Amsterdam, The Netherlands, 2018.
58. Roth, M.; Antwi, Y.A.; O'Sullivan, R. *Land and Natural Resource Governance and Tenure for Enabling Sustainable Cocoa Cultivation in Ghana*; USAID Tenure and Global Climate Change Program: Washington, DC, USA, 2017.
59. Ploton, P.; Mortier, F.; Réjou-Méchain, M.; Barbier, N.; Picard, N.; Rossi, V.; Dormann, C.; Cornu, G.; Viennois, G.; Bayol, N.; et al. Spatial Validation Reveals Poor Predictive Performance of Large-Scale Ecological Mapping Models. *Nat. Commun.* **2020**, *11*, 4540. [[CrossRef](#)] [[PubMed](#)]
60. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography* **2017**, *40*, 913–929. [[CrossRef](#)]
61. ESA. *Sentinel-2 User Handbook*, 2nd ed.; European Space Agency: Paris, France, 2015.
62. Simonetti, D.; Pimple, U.; Langner, A.; Marelli, A. Pan-Tropical Sentinel-2 Cloud-Free Annual Composite Datasets. *Data Brief* **2021**, *39*, 107488. [[CrossRef](#)]
63. Zupanc, A. Improving Cloud Detection with Machine Learning. Available online: <https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13> (accessed on 15 September 2023).
64. Google Earth Engine. Sentinel-2: Cloud Probability. Available online: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_CLOUD_PROBABILITY (accessed on 15 September 2023).
65. Skakun, S.; Wevers, J.; Brockmann, C.; Doxani, G.; Aleksandrov, M.; Batič, M.; Frantz, D.; Gascon, F.; Gómez-Chova, L.; Hagolle, O.; et al. Cloud Mask Intercomparison eXercise (CMIX): An Evaluation of Cloud Masking Algorithms for Landsat 8 and Sentinel-2. *Remote Sens. Environ.* **2022**, *274*, 112990. [[CrossRef](#)]
66. The Earth Engine Community Authors. Sentinel-2 Cloud Masking with s2cloudless. Available online: <https://github.com/google/earthengine-community/blob/master/tutorials/sentinel-2-s2cloudless/index.ipynb> (accessed on 15 September 2023).
67. Batista, J.E.; Rodrigues, N.M.; Cabral, A.I.R.; Vasconcelos, M.J.P.; Venturieri, A.; Silva, L.G.T.; Silva, S. Optical Time Series for the Separation of Land Cover Types with Similar Spectral Signatures: Cocoa Agroforest and Forest. *Int. J. Remote Sens.* **2022**, *43*, 3298–3319. [[CrossRef](#)]
68. Aguilar, R.; Zurita-Milla, R.; Izquierdo-Verdiguier, E.; De By, R.A. A Cloud-Based Multi-Temporal Ensemble Classifier to Map Smallholder Farming Systems. *Remote Sens.* **2018**, *10*, 729. [[CrossRef](#)]
69. Mishra, V.N.; Prasad, R.; Rai, P.K.; Vishwakarma, A.K.; Arora, A. Performance Evaluation of Textural Features in Improving Land Use/Land Cover Classification Accuracy of Heterogeneous Landscape Using Multi-Sensor Remote Sensing Data. *Earth Sci. Inform.* **2019**, *12*, 71–86. [[CrossRef](#)]
70. Sun, C.; Bian, Y.; Zhou, T.; Pan, J. Using of Multi-Source and Multi-Temporal Remote Sensing Data Improves Crop-Type Mapping in the Subtropical Agriculture Region. *Sensors* **2019**, *19*, 2401. [[CrossRef](#)] [[PubMed](#)]
71. Tavares, P.A.; Beltrão, N.E.; Guimarães, U.S.; Teodoro, A.C. Integration of Sentinel-1 and Sentinel-2 for Classification and LULC Mapping in the Urban Area of Belém, Eastern Brazilian Amazon. *Sensors* **2019**, *19*, 1140. [[CrossRef](#)] [[PubMed](#)]
72. Clevers, J.G.P.W.; Gitelson, A.A. Remote Estimation of Crop and Grass Chlorophyll and Nitrogen Content Using Red-Edge Bands on Sentinel-2 and -3. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 344–351. [[CrossRef](#)]
73. Hall-Beyer, M. Practical Guidelines for Choosing GLCM Textures to Use in Landscape Classification Tasks over a Range of Moderate Spatial Scales. *Int. J. Remote Sens.* **2017**, *38*, 1312–1338. [[CrossRef](#)]
74. Haralick, R.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *3*, 610–621. [[CrossRef](#)]
75. Huete, A.R.; Liu, H.Q.; Batchily, K.; Van Leeuwen, W. A Comparison of Vegetation Indices over a Global Set of TM Images for EOS-MODIS. *Remote Sens. Environ.* **1997**, *59*, 440–451. [[CrossRef](#)]
76. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
77. Gao, B.-C. NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [[CrossRef](#)]
78. Barnes, E.M.; Clarke, T.; Richards, S.; Colaizzi, P.D.; Haberland, J.; Kostrzewski, M.; Waller, P.; Choi, C.; Riley, E.; Thompson, T. Coincident Detection of Crop Water Stress, Nitrogen Status and Canopy Density Using Ground Based Multispectral Data. In Proceedings of the Fifth International Conference on Precision Agriculture, Bloomington, MN, USA, 16–19 July 2000.
79. Boiarskii, B.; Hasegawa, H. Comparison of NDVI and NDRE Indices to Detect Differences in Vegetation and Chlorophyll Content. *J. Mech. Contin. Math. Sci.* **2019**, *4*, 20–29. [[CrossRef](#)]
80. Rouse, J.W., Jr.; Haas, R.H.; Deering, D.; Schell, J.; Harlan, J.C. *Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation*; NASA Goddard Space Flight Center: Greenbelt, MD, USA, 1974; p. 371.
81. Merzlyak, M.N.; Gitelson, A.A.; Chivkunova, O.B.; Rakitin, V.Y. Non-Destructive Optical Detection of Pigment Changes during Leaf Senescence and Fruit Ripening. *Physiol. Plant.* **1999**, *106*, 135–141. [[CrossRef](#)]
82. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between Leaf Chlorophyll Content and Spectral Reflectance and Algorithms for Non-Destructive Chlorophyll Assessment in Higher Plant Leaves. *J. Plant Physiol.* **2003**, *160*, 271–282. [[CrossRef](#)]

83. Gitelson, A.A.; Merzlyak, M.N. Spectral Reflectance Changes Associated with Autumn Senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *J. Plant Physiol.* **1994**, *143*, 286–292. [CrossRef]
84. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]
85. Haboudane, D.; Tremblay, N.; Miller, J. Remote Estimation of Crop Chlorophyll Content Using Spectral Indices Derived from Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 423–437. [CrossRef]
86. Gitelson, A.A.; Kaufman, Y.; Stark, R.; Rundquist, D. Novel Algorithms for Remote Estimation of Vegetation Fraction. *Remote Sens. Environ.* **2002**, *80*, 76–87. [CrossRef]
87. Connors, R.W.; Trivedi, M.M.; Harlow, C.A. Segmentation of a High-Resolution Urban Scene Using Texture Operators. *Comput. Vis. Graph. Image Process.* **1984**, *25*, 273–310. [CrossRef]
88. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. [CrossRef]
89. Google Colaboratory Team. Colaboratory. Available online: <https://workspace.google.com/marketplace/app/colaboratory/1014160490159> (accessed on 17 September 2023).
90. Müllner, D. Modern Hierarchical, Agglomerative Clustering Algorithms. *arXiv* **2011**, arXiv:1109.2378. [CrossRef]
91. Ward, J.H., Jr. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244. [CrossRef]
92. Cunningham, P.; Kathirgamanathan, B.; Delany, S.J. Feature Selection Tutorial with Python Examples. *arXiv* **2021**, arXiv:2106.06437. [CrossRef]
93. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.; Tang, J.; Liu, H. Feature Selection: A Data Perspective. *ACM Comput. Surv.* **2017**, *50*, 1–45. [CrossRef]
94. Gregorutti, B.; Michel, B.; Saint-Pierre, P. Correlation and Variable Importance in Random Forests. *Stat. Comput.* **2017**, *27*, 659–678. [CrossRef]
95. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
96. Strobl, C.; Boulesteix, A.L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional Variable Importance for Random Forests. *BMC Bioinform.* **2008**, *9*, 307. [CrossRef]
97. Gómez-Ramírez, J.; Ávila-Villanueva, M.; Fernández-Blázquez, M.Á. Selecting the Most Important Self-Assessed Features for Predicting Conversion to Mild Cognitive Impairment with Random Forest and Permutation-Based Methods. *Sci. Rep.* **2020**, *10*, 20630. [CrossRef]
98. Belgiu, M.; Drăguț, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
99. Deng, C.; Wu, C. The Use of Single-Date MODIS Imagery for Estimating Large-Scale Urban Impervious Surface Fraction with Spectral Mixture Analysis and Machine Learning Techniques. *ISPRS J. Photogramm. Remote Sens.* **2013**, *86*, 100–110. [CrossRef]
100. Du, P.; Samat, A.; Waske, B.; Liu, S.; Li, Z. Random Forest and Rotation Forest for Fully Polarized SAR Image Classification Using Polarimetric and Spatial Features. *ISPRS J. Photogramm. Remote Sens.* **2015**, *105*, 38–53. [CrossRef]
101. Ghimire, B.; Rogan, J.; Galiano, V.R.; Panday, P.; Neeti, N. An Evaluation of Bagging, Boosting, and Random Forests for Land-Cover Classification in Cape Cod, Massachusetts, USA. *GISci. Remote Sens.* **2012**, *49*, 623–643. [CrossRef]
102. Pelletier, C.; Valero, S.; Inglada, J.; Champion, N.; Dedieu, G. Assessing the Robustness of Random Forests to Map Land Cover with High Resolution Satellite Image Time Series over Large Areas. *Remote Sens. Environ.* **2016**, *187*, 156–168. [CrossRef]
103. Condro, A.A.; Setiawan, Y.; Prasetyo, L.B.; Pramulya, R.; Siahaan, L. Retrieving the National Main Commodity Maps in Indonesia Based on High-Resolution Remotely Sensed Data Using Cloud Computing Platform. *Land* **2020**, *9*, 377. [CrossRef]
104. Chan, J.C.-W.; Beckers, P.; Spanhove, T.; Borre, J.V. An Evaluation of Ensemble Classifiers for Mapping Natura 2000 Heathland in Belgium Using Spaceborne Angular Hyperspectral (CHRIS/Proba) Imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 13–22. [CrossRef]
105. Erinjery, J.J.; Singh, M.; Kent, R. Mapping and Assessment of Vegetation Types in the Tropical Rainforests of the Western Ghats Using Multispectral Sentinel-2 and SAR Sentinel-1 Satellite Imagery. *Remote Sens. Environ.* **2018**, *216*, 345–354. [CrossRef]
106. Lawrence, R.L.; Moran, C.J. The AmericaView Classification Methods Accuracy Comparison Project: A Rigorous Approach for Model Selection. *Remote Sens. Environ.* **2015**, *170*, 115–120. [CrossRef]
107. Pelletier, C.; Webb, G.L.; Petitjean, F. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sens.* **2019**, *11*, 523. [CrossRef]
108. Vaglio Laurin, G.; Liesenberg, V.; Chen, Q.; Guerriero, L.; Del Frate, F.; Bartolini, A.; Coomes, D.; Wilebore, B.; Lindsell, J.; Valentini, R. Optical and SAR Sensor Synergies for Forest and Land Cover Mapping in a Tropical Site in West Africa. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 7–16. [CrossRef]
109. Congalton, R.G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [CrossRef]
110. Pontius, R.G.; Millones, M. Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429. [CrossRef]
111. Stehman, S.V. Selecting and Interpreting Measures of Thematic Classification Accuracy. *Remote Sens. Environ.* **1997**, *62*, 77–89. [CrossRef]

112. Inglada, J.; Arias, M.; Tardy, B.; Hagolle, O.; Valero, S.; Morin, D.; Dedieu, G.; Sepulcre, G.; Bontemps, S.; Defourny, P.; et al. Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote Sens.* **2015**, *7*, 12356–12379. [[CrossRef](#)]
113. Inglada, J.; Vincent, A.; Arias, M.; Tardy, B.; Morin, D.; Rodes, I. Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sens.* **2017**, *9*, 95. [[CrossRef](#)]
114. Lyons, M.B.; Keith, D.A.; Phinn, S.R.; Mason, T.J.; Elith, J. A Comparison of Resampling Methods for Remote Sensing Classification and Accuracy Assessment. *Remote Sens. Environ.* **2018**, *208*, 145–153. [[CrossRef](#)]
115. Google Earth. Google Earth Versions. Available online: <https://www.google.com/intl/en/earth/versions/> (accessed on 17 September 2023).
116. QGIS.org. QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available online: <http://qgis.org> (accessed on 17 September 2023).
117. Hermosilla, T.; Wulder, M.A.; White, J.C.; Coops, N.C. Land Cover Classification in an Era of Big and Open Data: Optimizing Localized Implementation and Training Data Selection to Improve Mapping Outcomes. *Remote Sens. Environ.* **2022**, *268*, 112780. [[CrossRef](#)]
118. Pelletier, C.; Valero, S.; Inglada, J.; Champion, N.; Marais Sicre, C.; Dedieu, G. Effect of Training Class Label Noise on Classification Performances for Land Cover Mapping with Satellite Image Time Series. *Remote Sens.* **2017**, *9*, 173. [[CrossRef](#)]
119. Qin, R.; Liu, T. A Review of Landcover Classification with Very-High Resolution Remotely Sensed Optical Images—Analysis Unit, Model Scalability and Transferability. *Remote Sens.* **2022**, *14*, 646. [[CrossRef](#)]
120. Anyimah, F.O.; Osei, E.M., Jr.; Nyamekye, C. Detection of Stress Areas in Cocoa Farms Using GIS and Remote Sensing: A Case Study of Offinso Municipal & Offinso North District, Ghana. *Environ. Chall.* **2021**, *4*, 100087. [[CrossRef](#)]
121. Millard, K.; Richardson, M. On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sens.* **2015**, *7*, 8489–8515. [[CrossRef](#)]
122. Chen, D.; Stow, D.A.; Gong, P. Examining the Effect of Spatial Resolution and Texture Window Size on Classification Accuracy: An Urban Environment Case. *Int. J. Remote Sens.* **2004**, *25*, 2177–2192. [[CrossRef](#)]
123. Hall-Beyer, M. *GLCM Texture: A Tutorial*, 3rd ed.; Department of Geography, University of Calgary: Calgary, AB, Canada, 2017.
124. Roberts, D.; Mueller, N.; Mcintyre, A. High-Dimensional Pixel Composites from Earth Observation Time Series. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6254–6264. [[CrossRef](#)]
125. Bey, A.; Jetimane, J.; Lisboa, S.N.; Ribeiro, N.; Siteo, A.; Meyfroidt, P. Mapping Smallholder and Large-Scale Cropland Dynamics with a Flexible Classification System and Pixel-Based Composites in an Emerging Frontier of Mozambique. *Remote Sens. Environ.* **2020**, *239*, 111611. [[CrossRef](#)]
126. Li, Q.; Qiu, C.; Ma, L.; Schmitt, M.; Zhu, X.X. Mapping the Land Cover of Africa at 10 m Resolution from Multi-Source Remote Sensing Data with Google Earth Engine. *Remote Sens.* **2020**, *12*, 602. [[CrossRef](#)]
127. Scott, G.J.; England, M.; Starms, W.A.; Marcum, R.A.; Davis, C.H. Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 549–553. [[CrossRef](#)]
128. Tuia, D.; Volpi, M.; Copa, L.; Kanevski, M.F.; Muñoz-Marí, J. A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 606–617. [[CrossRef](#)]
129. Turkoglu, M.O.; D’Aronco, S.; Perich, G.; Liebisch, F.; Streit, C.; Schindler, K.; Wegner, J.D. Crop Mapping from Image Time Series: Deep Learning with Multi-Scale Label Hierarchies. *Remote Sens. Environ.* **2021**, *264*, 112603. [[CrossRef](#)]
130. Crawford, M.M.; Tuia, D.; Yang, H.L. Active Learning: Any Value for Classification of Remotely Sensed Data? *Proc. IEEE* **2013**, *101*, 593–608. [[CrossRef](#)]
131. Critchley, M.; Sassen, M.; Umunay, P. *Mapping Opportunities for Cocoa Agroforestry in Côte d’Ivoire: Assessing its Potential to Contribute to National Forest Cover Restoration Targets and Ecosystem Services Co-Benefits*; UNEP World Conservation Monitoring Centre: Cambridge, UK, 2021.
132. Mighty Earth. *Sweet Nothings: How the Chocolate Industry Has Failed to Honor Promises to End Deforestation for Cocoa in Côte d’Ivoire and Ghana*; Mighty Earth: Washington, DC, USA, 2022.
133. Copernicus. Copernicus HotSpot Land Cover Change Explorer. Available online: <https://land.copernicus.eu/global/hsm> (accessed on 27 September 2023).

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