







Article

Integrating the Strength of Multi-Date Sentinel-1 and -2 Datasets for Detecting Mango (*Mangifera indica* L.) Orchards in a Semi-Arid Environment in Zimbabwe

Bester Tawona Mudereri ^{1,*}, Elfatih M. Abdel-Rahman ¹, Shepard Ndlela ¹, Louisa Delfin Mutsa Makumbe ², Christabel Chiedza Nyanga ², Henri E. Z. Tonnang ¹ and Samira A. Mohamed ¹

¹ International Centre of Insect Physiology and Ecology (ICIPE), P.O. Box 30772, 00100 Nairobi, Kenya; eabdel-rahman@icipe.org (E.M.A.-R.); sndlela@icipe.org (S.N.); htonnang@icipe.org (H.E.Z.T.); sfaris@icipe.org (S.A.M.)

² Plant Quarantine Services Institute, 33 km peg Mazowe Bindura Highway, P. Bag 2007, Mazowe, Zimbabwe; makumbelouisa@gmail.com (L.D.M.M.); chiedzanyanga@gmail.com (C.C.N.)

* Correspondence: bmudereri@icipe.org

Abstract: Generating tree-specific crop maps within heterogeneous landscapes requires imagery of fine spatial and temporal resolutions to discriminate among the rapid transitions in tree phenological and spectral features. The availability of freely accessible satellite data of relatively high spatial and temporal resolutions offers an unprecedented opportunity for wide-area land use and land cover (LULC) mapping, including tree crop (e.g., mango; *Mangifera indica* L.) detection. We evaluated the utility of combining Sentinel-1 (S1) and Sentinel-2 (S2) derived variables ($n = 81$) for mapping mango orchard occurrence in Zimbabwe using machine learning classifiers, i.e., support vector machine and random forest. Field data were collected on mango orchards and other LULC classes. Fewer variables were selected from 'All' combined S1 and S2 variables using three commonly utilized variable selection methods, i.e., relief filter, guided regularized random forest, and variance inflation factor. Several classification experiments ($n = 8$) were conducted using 60% of field datasets and combinations of 'All' and fewer selected variables and were compared using the remaining 40% of the field dataset and the area underclass approach. The results showed that a combination of random forest and relief filter selected variables outperformed (F1 score > 70%) all other variable combination experiments. Notwithstanding, the differences among the mapping results were not significant ($p \leq 0.05$). Specifically, the mapping accuracy of the mango orchards was more than 80% for each of the eight classification experiments. Results revealed that mango orchards occupied approximately 18% of the spatial extent of the study area. The S1 variables were constantly selected compared with the S2-derived variables across the three variable selection approaches used in this study. It is concluded that the use of multi-modal satellite imagery and robust machine learning classifiers can accurately detect mango orchards and other LULC classes in semi-arid environments. The results can be used for guiding and upscaling biological control options for managing mango insect pests such as the devastating invasive fruit fly *Bactrocera dorsalis* (Hendel) (Diptera: *Tephritidae*).

Keywords: Google Earth Engine (GEE); land use and land cover (LULC); machine learning; multi-algorithm classification; pest control; variable selection; SAR



Citation: Mudereri, B.T.; Abdel-Rahman, E.M.; Ndlela, S.; Makumbe, L.D.M.; Nyanga, C.C.; Tonnang, H.E.Z.; Mohamed, S.A. Integrating the Strength of Multi-Date Sentinel-1 and -2 Datasets for Detecting Mango (*Mangifera indica* L.) Orchards in a Semi-Arid Environment in Zimbabwe. *Sustainability* **2022**, *14*, 5741. <https://doi.org/10.3390/su14105741>

Academic Editor: José Manuel Mirás-Avalos

Received: 21 March 2022

Accepted: 28 April 2022

Published: 10 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Spatial characterization and mapping of land use and land cover (LULC) provide crucial planning tools and decision support systems through the quantification and evaluation of the spatial distribution of fundamental natural resources such as mango (*Mangifera indica* L.) orchards [1–4]. Such tools can be used for numerous strategies, such as the implementation and upscaling of integrated pest management (IPM) options, estimation of crop

production, and assessing the distribution and resilience of food systems. Additionally, LULC features can be utilized as primary input variables in the geospatial crop, pest, and pollinator modeling approaches [5]. The LULC information is thus key in strengthening the spatio-temporal agricultural planning and driving progress towards achieving the overarching United Nations sustainable development goals (UN SDGs) and commitment to end hunger, food insecurity, and all forms of malnutrition (availability, access, and utilization) by 2030 [6]. Mango is among the crop systems that can substantially contribute to achieving the mentioned SDGs by providing food and generating income for impoverished smallholder farmers [7,8].

Worldwide, mango is one of the most important nutritional and cash tree fruit crops in tropical and subtropical regions [9]. The crop provides income and sustains food and nutritional security for approximately 25% of the smallholder farmers in sub-Saharan Africa [8]. In 2020, the mango-harvested area in Africa was $\sim 967 \times 10^3$ ha, contributing to *ca* 12.3% of the global mango production [10]. The demand for mangos and other various fruits and vegetables is increasing globally because of population growth, societal affluence, improved lifestyles, and the general increase in health awareness and nutritive benefits [8,11]. This demand is also attributed to the 400 g recommended daily intake of fruit and vegetables by the food and agriculture organization (FAO) and the world health organization (WHO) [8,11]. Avocado, mango, orange, peach, and guava were identified as the most popular exotic fruit tree species commonly consumed in Zimbabwe, with an average mango daily consumption of about 40 g/day/person [12]. Therefore, the determination of these crucial fruit tree area coverage in Africa is critical for assessing their occurrence, abundance, yield, production, and post-harvest treatments [10]. Moreover, delineating mango areal extent is important for other inventory and management information, such as crop insurance and the monitoring of biotic and abiotic stresses. For instance, spatio-temporal changes in mango systems could be related to climate shocks and insect pest (e.g., fruit flies) damage [13–15].

There are several available remotely-sensed LULC products with medium to high spatial resolutions at a global scale that could be used for mango monitoring and management. For example, the European space agency (ESA) GlobCover (300 m \times 300 m pixel size), Copernicus global land cover (100 m \times 100 m pixel size), Globland (30 m \times 30 m pixel size), FROM-GLC (10–30 m \times 10–30 m pixel size), and the 20 m \times 20 m pixel size ESA climate change initiative (CCI) land cover prototype. However, all these readily-available products fail to provide explicit tree species-specific maps. Additionally, the accuracy of these products is comparatively low, with many discrepancies existing among them. As such, there is a limit on the quality of the information necessary to improve planning initiatives in sub-Saharan Africa regarding mango orchard mapping. Thus, the insufficiency of the existing LULC products for localized or species-specific cover mapping highlights the need for developing innovative methods that provide more efficient, accurate, and reliable LULC maps.

The development of such methods is achievable because of the advancement and improved accessibility to freely available satellite imageries of medium spatial and temporal resolutions that open new prospects for LULC mapping including tree-specific orchards detection [16]. For instance, remote sensing satellite-based observations such as the constellation of Sentinel-2 (S2) are a distinctive source of data for LULC characterization because of their near-real-time precision, worldwide coverage, repetitiveness, and dependability at a free cost [17]. The S2 satellite imageries have been tested by several studies and recommended for use in many environmental monitoring applications that include identification of tree species [18], rangeland quality evaluation [19], coffee crop detection [20], cropland mapping [21], and general LULC mapping [22,23]. The constellation of S2 includes the identical optical S2A and S2B satellites launched in June 2015 and March 2017, respectively, offering a wider swath (290 km), medium spatial resolution (10, 20, and 60 m), high temporal resolution (5 days revisit time for S2A and B combined), and multi-spectral (13 spectral bands) capabilities [24,25]. Amongst the S2 13 spectral bands, there are four

unique and useful red-edge bands (bands 5, 6, 7, and 8A) with central wavebands of 704.1 nm, 740.5 nm, 782.8 nm, and 864.7 nm, respectively, that enhance the possibility of class-specific separability among LULC classes that exhibit similar spectral characteristics, such as croplands and grasslands [26].

Although S2 imagery provides the above-mentioned momentous advantages over other optical satellite imagery such as Landsat and moderate resolution imaging spectroradiometer (MODIS), all these datasets including S2 are often affected by the time of day (night) and weather patterns such as clouds and inter-seasonal differences [26]. However, the high temporal resolution of the S2 constellation (S2A and S2B) offers the possibility to harness intra- and inter-seasonal reflectance changes to identify earth objects over time [23]. This advantage is particularly essential in regions with strong vegetation seasonality, such as the semi-arid tropics with distinct wet and dry seasons [27]. Additionally, the sentinel mission offers observations using a Sentinel-1 (S1) sensor that operates in C-band synthetic aperture radar (SAR). The benefit of such a SAR band is that it is not affected by cloud cover or insufficient illumination and acquires data under all weather conditions during day or night time [16,28]. Hence, the sentinel mission provides an opportunity for combining both S2 optical and S1 SAR data to enhance the mapping capacity in situations where multi-date optical satellite-based data are limited due to unfavorable imaging conditions. Recent studies have successfully combined the advantages of both S2 and S1 to detect and differentiate LULC classes in various agro-natural ecosystems [23,29,30]. However, accurate and reliable LULC mapping is not possible without employing advanced and efficient machine learning (ML) classifiers [31,32]. Thus, the appropriate remotely sensed and associated ancillary data sources, optimum precisions, and the finest classification techniques for LULC mapping are still debatable within the remote sensing community [2].

Several ML classifiers are comparatively better than the conventional statistical classifiers in producing LULC maps with varying levels of performance. Commonly, the performance variation among ML algorithms in classifying LULC classes is caused using different combinations of predictor remotely-sensed variables, sampling strategies used in the collection of georeferenced points, and the complexity of the agroecology in different geographic regions [2]. Additionally, the strength and capability of the ML classifier itself used for processing, mining, and analyzing the predictor variables with the classes can play a major role in obtaining varying LULC mapping results. Thus, it is beneficial to take advantage and test the strengths of different ML classifiers for mapping tree-specific fruit crops. As mentioned earlier, accurate and reliable tree-specific mapping results should also be coupled with the incorporation of multiple satellite-based data from various sensor (e.g., S1 and S2) types [23]. Notwithstanding, multiple datasets with numerous predictor variables might be highly dimensional, redundant, and multicollinear. This could cause an increase in computational density and time, and ingest more computer storage space [33]. Moreover, the inclusion of highly correlated predictor variables in a classification experiment can make the classification model very sensitive to any minor change in the model settings and parameters. A highly sensitive classification model reduces the precision and accuracy of the predictions by weakening the computational power of the model. Hence, there is a need for applying variable selection techniques before performing a classifier [33].

One of the ways to optimize the selection of relevant predictor variables is to apply data exploration procedures that are aimed at obtaining fewer computationally efficient variables that are sensitive to relatively strong patterns of association (i.e., interactions) so that informative variables are not erroneously eliminated before executing the classification experiment [34]. In general, the currently available variable selection techniques fall within three main categories [35,36]: (i) wrapper methods (e.g., forward, backward, and stepwise selection approaches), (ii) filter methods (e.g., ANOVA, correlation, variance thresholding), and (iii) embedded methods (e.g., lasso, ridge, decision trees), which have been successfully used in earlier studies that utilized remotely sensed datasets [37]. Unlike the other methods, the embedded ones are essentially ML algorithms, among which Lasso [38], random forest (RF: [39]), and the guided regularized random forest (GRRF: [40]) are the most powerful and widely used in analyzing remotely sensed data.

In the literature, few or no studies have attempted to combine the strengths of SAR and optical datasets to reveal the best predictor combinations derived using embedded variable selection methods and ML classifiers for mango orchard detection. Furthermore, the performance of RF and support vector machines (SVM: [41]) classifiers have never been compared to detect mango orchards within a complex semi-arid environment. Therefore, herein, we tested the strength of S1 and S2 as SAR and optical sensors, respectively, for mango orchard detection in a semi-arid environment in Zimbabwe. In particular, we examined the utility of the S1 backscatter and S2 reflectance, derived vegetation indices, and three variable selection methods (relief filter (relieff), GRRF, and variance inflation factor (VIF)) to discriminate between eight LULC classes including mango orchard using the RF and SVM machine learning classification algorithms, respectively. Furthermore, our study established the most relevant SAR and optical remotely sensed variables necessary for mapping mango orchards and other LULC classes in semi-arid environments.

2. Study Area

The study was conducted in the Mashonaland East (Mutoko and Murehwa districts) and Mashonaland West (Zvimba district) provinces of Zimbabwe. The administrative locations of the three districts are at Mutoko ($17^{\circ}10'0''$ S; $32^{\circ}30'0''$ E), Murehwa ($17^{\circ}48'0''$ S; $31^{\circ}50'0''$ E), and Zvimba ($17^{\circ}42'0''$ S; $30^{\circ}12'0''$ E) (Figure 1). A unimodal rainfall pattern characterizes most of the agro-ecological regions in southern Africa, including Zimbabwe [42]. In the study area districts, annual rainfall ranges between 750 and 1000 mm for Murehwa and Zvimba and between 650 and 750 mm for the Mutoko district. The temperature in winter has a minimum average of 10°C , while summer temperatures can reach a maximum average of 33°C across the three districts [43]. Isolated woody vegetation, bushland, grasslands, and rocky hills dominate the ecosystems in the study area. Most of the communal farmers perform subsistence agriculture as the main economic activity [44]. Maize is the major farming system in both the communal and commercial farming systems in the three districts. Additionally, the three districts are some of the most dominant mango production areas in Zimbabwe, hence our motivation for selecting them as our study sites [12].

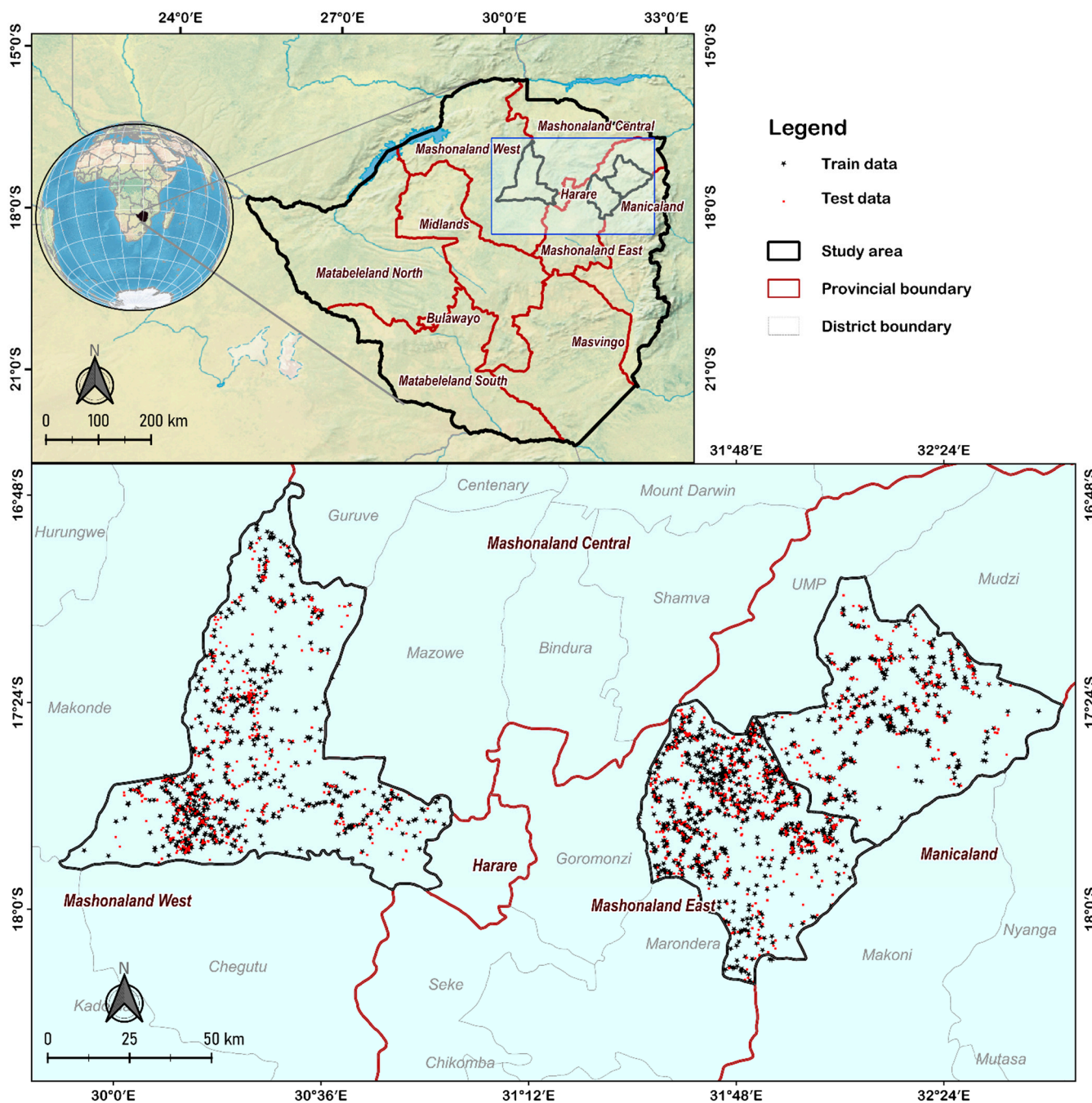


Figure 1. Location of the study area in Africa and Zimbabwe along with the spatial distribution of land use and land cover (LULC) ground reference points taken between the 23 May 2021 and the 4 June 2021 superimposed on shaded relief natural Earth data (<https://www.naturalearthdata.com/>).

3. Methodology

Our proposed semi-automated mango mapping methodology uses two satellite-based data types derived from SAR (i.e., S1) and optical (i.e., S2) sensors, together with three variable selection methods and two ML classification algorithms to develop explicit LULC maps in Google Earth Engine (GEE). The generalized workflow of our semi-automated mango orchards classification approach used in this study is shown in Figure 2. The specifics of every procedure are explained in detail in the subsequent subsections.

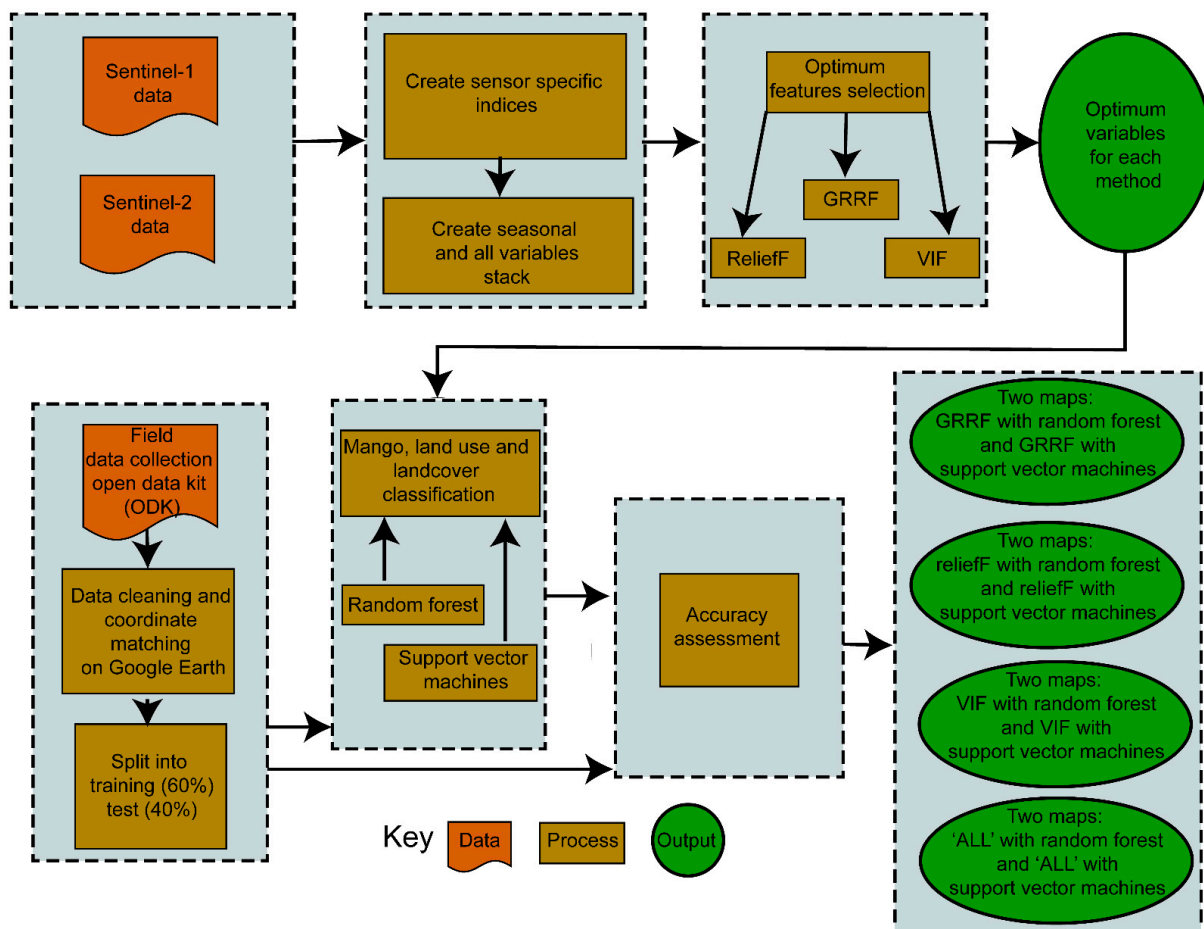


Figure 2. The general workflow of the methodology used to test the strength of Sentinel-1 and -2 for mapping mango orchards, together with other land use and land cover (LULC) classes in a semi-arid environment in Zimbabwe.

3.1. Land Use and Land Cover (LULC) Classes

A total of eight LULC classes were used to characterize the landscape in the study area. These classes were informed by the general structure of semi-arid environments, the specific structure of our study area, and the importance of the classes in assessing the biotic and abiotic risk factors for agricultural production. Specifically, the eight classes comprised mango, bare soil, urban/built-up, cropland, forest, grassland, shrubland, and water. The mango class was the main target class that identified mango trees occurring at homesteads, within fields, or in the wild at a minimum tree height of >1.7 m and a canopy cover of 225 m² (15 × 15 m). All juvenile mango shrubs below the mentioned height and canopy coverage were not considered. We opted to only include mango orchards of 225 m² to ensure that an orchard sample would at least match up with the Sentinel pixel size, viz., 10 m × 10 m. Hence, individually occurring mango trees were excluded from the field data collection. The ‘bare soil’ class referred to open areas or areas with scant grass vegetation of up to 10% or rocky areas, e.g., a Dwala. Urban/built-up areas included settlements, i.e., houses and impervious, paved, and unpaved surfaces. Cropland referred to planted crop areas of maize, millets, sorghum, cowpeas, vegetables, etc. Those areas that were considered as the cropland class either had standing crops or there was evidence of recent harvest when the LULC data were collected. Therefore, abandoned croplands were considered as either grassland or shrubland based on their constituents. Forests were identified as an assemblage of trees of greater than 2 m in height with some understory, while ‘grasslands’ were devoid of trees and mainly dominated by grasses. Shrublands were areas with sparse, short bushes of heights less than 2 m and with coppicing covering

between 20% and 80% of the soil surface within 225 m² (15 m × 15 m). The ‘water’ class referred to permanent open water bodies such as rivers and lakes only.

3.2. Mango and Other Land Use and Land Cover (LULC) Field Data Collection

Mango tree orchard locations ($n = 1072$), together with other LULC-class ($n = 1931$) data, were obtained during field surveys conducted between 23 May 2021 and 4 June 2021, following a stratified random sampling protocol. We stratified the study area using the administrative wards as strata (29 in Mutoko, 30 in Murehwa, and 35 in Zvimba) and randomly collected the LULC classes in each stratum (i.e., ward). Because the main aim was to map mango orchards in the study area, about 35% of our LULC samples were on mango orchards (Table 1). Herein, we leveraged the strength and capabilities of the open data kit (ODK) smartphone application (*app*), and the vast network of agricultural extension services personnel in Zimbabwe to capture the location of the eight thematic classes used in the LULC classification. The ODK is a multi-operational *app* parameterized and optimized to efficiently assemble, aggregate, and analyze survey data [45]. Furthermore, the ODK provides a compound *app* logic and strengthens the possibility of data manipulation that comprises text, geographic position, images, audio, video, and barcodes [46]. An optimized questionnaire to capture the necessary and required LULC information was created in the ‘ODK Build’ and was successively installed in the ‘ODK Collect’ module available on ‘Android’ devices of technically-trained agricultural extension officers. Thereafter, the collected observations on LULC were accessed, downloaded, and visualized. Additionally, the global positioning system (GPS) capabilities within the ODK, which provide ± 4 m accuracy, was utilized to geolocate the LULC observations in the field. A total of 3117 sample points were originally sent to the ODK server. These data were put through a rigorous data cleaning mechanism to eliminate duplicates and misplaced coordinates using the R software (version 4.0.4, Vienna, Austria) [47]. The misplaced coordinates were verified and eliminated from the data set using the Google Earth platform (<https://earth.google.com/>) as a reference. After the data cleaning process, a combined total of 3003 reference points were retained for further analysis. We utilized the ‘caret’ package [48] in R software to split the reference LULC observation points for each class using a ratio of 60% for training and 40% for testing the accuracy of our classification algorithms (Table 1). The locations and distribution of both the training and testing LULC datasets used in this study are shown in Figure 1. The independently split ground reference points were converted to shapefiles and uploaded into the GEE for further analysis.

Table 1. Mango and other land use and land cover (LULC) classes, the sample size of the training and test datasets that were used in the classification experiment of mapping mango orchards and other LULC classes in the Mutoko, Murehwa, and Zvimba districts in Zimbabwe.

LULC Class	Class Number	Training Data (60%)	Test Data (40%)	Total
Mango	1	644	428	1072
Built-up	2	227	150	377
Cropland	3	263	175	438
Forest	4	173	114	287
Grassland	5	185	122	307
Bare soil	6	83	55	138
Shrubland	7	88	58	146
Water	8	143	95	238
Total		1806	1197	3003

3.3. Sentinel-2 Image Processing in Google Earth Engine (GEE)

The GEE platform offers parallelized processing steps on the Google cloud that enables the processing and analysis of satellite imagery at a petabyte-scale [30,49]. The platform includes freely available various satellite data such as Landsat, S1, and S2, among others [30,50]. Additionally, GEE provides various advanced ML analytics and ancillary data types, such as bioclimatic, and remotely sensed vegetation indicators [50]. In this study, the S2 image preprocessing steps were automatically conducted in GEE following the procedures suggested by Schulz et al. [23]. The S2 imageries in the GEE were provided by the Copernicus program of the European Space Agency [24] as level 2A. Level 2A products of S2 provide bottom of atmosphere (BOA) reflectance images that were derived from the associated level-1C products. Specifically, each S2 level-2A product is composed of 12 bands (excluding band 10) as 100 km × 100 km tiles in the universal transverse mercator (UTM, zone 36 south) and WGS84 projection [24]. We utilized multi-date S1 ($n = 50$) and S2 ($n = 60$) imageries that were subjected to several preprocessing steps. The selection of the multi-date imagery was guided by the cropping seasons between 2020 and 2021 (i.e., 2020/2021 cropping calendar) in Zimbabwe and by cloud cover in the imagery ($\leq 30\%$). The 2020/2021 cropping calendar was selected to guide the selection of imagery dates as it was in general regarded as a normal rainfall year with no distinct anomalies, and this provided complete season data [51]. In particular, the S1 and S2 imagery were acquired during the dry season, 1 June 2020–15 October 2020 ($n = 20$ for S1 and 24 for S2); harvest season, 1–30 May 2020 ($n = 5$ for S1 and 6 for S2); and wet season, 1 November 2020–30 April 2021 ($n = 25$ for S1 and 30 for S2) of Zimbabwe.

Our automated S2 image preprocessing steps in GEE included: (1) resampling of all S2 image bands that had pixel size greater than 10 m × 10 m (Bands 1, 5, 6, 7, 8A, 9, 11, and 12) to the lowest spatial resolution (10 m × 10 m) of S2 using the nearest neighbor approach [50]; (2) masking the imagery cloud and shadow using a filtering process of the scene classification (SC) band provided in the level 2A products; and (3) calculation of a median pixel value for each S2 band as per the cropping season [24]. The median pixel values for all S2 bands ($n = 12$) were then used as predictor variables or for calculating seven vegetation indices, which were also combined with the bands as predictor variables for mapping mango and other LULC classes (Table 2). Studies have shown that the median image compositing method offers relatively better results than other methods, such as a maximum ratio value [52,53]. Seven S2-based vegetation indices comprised the normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), normalized difference moisture index (NDMI), normalized difference built-up index (NDBI), and three tasseled cap (TC) transformed indices for brightness (TCBI), wetness index (TCWI), and greenness index (TCVI). Similarly, two predictor variables comprising the seasonal standard deviations of NDVI (NDVI_stdDev) and EVI (EVI_stdDev) were also computed. The syntax and Equations (1)–(7) of these selected indices were adapted from the index database (IDB) <https://www.indexdatabase.de/> [54]. The IDB is a tool that was developed to provide a simple overview of satellite-specific vegetation indices, which are usable from a specific sensor for a specific application [54]. The selection of these seven S2-based indices was motivated by the fact that they can efficiently capture the sensitivity of vegetation variables while minimizing the atmospheric and soil background noises on the image reflectance [55]. Additionally, these selected indices have been reported by other studies as the most suitable remotely sensed variables for capturing vegetation variabilities over time [56].

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{Red}}}{\rho_{\text{NIR}} + \rho_{\text{Red}}} \quad (1)$$

$$\text{EVI} = 2.5 \times \frac{\rho_{\text{NIR}} - \rho_{\text{Red}}}{\rho_{\text{NIR}} + 6 \times \rho_{\text{Red}} - 7.5 \times \rho_{\text{Blue}} + 1} \quad (2)$$

$$\text{NDMI} = \frac{\rho_{\text{NIR}} - \rho_{\text{SWIR1}}}{\rho_{\text{Red}} + \rho_{\text{SWIR1}}} \quad (3)$$

$$\text{NDBI} = \frac{\rho_{\text{SWIR1}} - \rho_{\text{NIR}}}{\rho_{\text{SWIR1}} + \rho_{\text{NIR}}} \quad (4)$$

$$\text{TCBI} = 0.3037 \rho_{\text{Blue}} + 0.2793 \rho_{\text{Green}} + 0.4743 \rho_{\text{Red}} + 0.5585 \rho_{\text{NIR}} + 0.5082 \rho_{\text{SWIR1}} + 0.1863 \rho_{\text{SWIR2}} \quad (5)$$

$$\text{TCWI} = 0.1509 \rho_{\text{Blue}} + 0.1973 \rho_{\text{Green}} + 0.3279 \rho_{\text{Red}} + 0.3406 \rho_{\text{NIR}} - 0.7112 \rho_{\text{SWIR1}} - 0.4572 \rho_{\text{SWIR2}} \quad (6)$$

$$\text{TCVI} = -0.2848 \rho_{\text{Blue}} - 0.243 \rho_{\text{Green}} - 0.5436 \rho_{\text{Red}} + 0.7243 \rho_{\text{NIR}} + 0.084 \rho_{\text{SWIR1}} - 0.18 \rho_{\text{SWIR2}} \quad (7)$$

where ρ_{Blue} (S2 band 2), ρ_{Green} (S2 band 3), ρ_{Red} (S2 band 4), ρ_{NIR} (S2 band 8), ρ_{SWIR1} (S2 band 11), and ρ_{SWIR2} (S2 band 12) in Equations (1)–(7) represent the blue, green, red, near-infrared, shortwave infrared 1, and shortwave infrared 2 reflectance values, respectively, for a given pixel.

Table 2. Summary description of the 21 Sentinel-2 derived bands and indices that were used across the three seasonal clusters, i.e., wet, dry, and harvest seasons. Band 10 (Cirrus) of Sentinel-2 is excluded in the level-2 processing. All bands were resampled from their original spatial resolution to $10 \text{ m} \times 10 \text{ m}$.

Band Number/Index	Description	Central Wavelength (nm)	Width	Original Spatial Resolution (m)
B1	Coastal aerosol	443	20	60
B2	Blue	490	65	10
B3	Green	560	35	10
B4	Red	665	30	10
B5	Red-edge (RE5)	705	15	20
B6	Red-edge (RE6)	740	15	20
B7	Red-edge (RE7)	783	20	20
B8	Near infrared	842	115	10
B8A	Rededge NIR	865	20	20
B9	Water vapor	945	20	60
B11	Short wave infrared	1610	90	20
B12	Short wave infrared	2190	180	20
EVI	Enhanced vegetation index	NA ¹	NA	NA
EVI_stdDev ²	EVI standard deviation	NA	NA	NA
NDBI	Normalized difference build-up index	NA	NA	NA
NDMI	Normalized difference moisture index	NA	NA	NA
NDVI	Normalized difference vegetation index	NA	NA	NA
NDVI_stdDev	NDVI standard deviation	NA	NA	NA
TCBI	Tasseled cap brightness index	NA	NA	NA
TCVI	Tasseled cap greenness index	NA	NA	NA
TCWI	Tasseled cap wetness index	NA	NA	NA

¹ NA = not applicable, ² stdDev = standard deviation.

3.4. Sentinel-1 (S1) Image Processing in Google Earth Engine (GEE)

The feasibility of the SAR data for mapping LULC has already been tested and established [28,57–59]. Studies have revealed that the interaction of SAR signal with the vegetation is volumetric and quite sensitive to the canopy structure, orientation, and moisture content [60]. Specifically, the variability in backscatter of, for instance, SAR vegetation, bare soil, water, and built-up signals can be uniquely different and distinguishable [60]. Moreover, SAR data complement the limitation of optical sensors in that their signals do not penetrate the clouds. Hence, in cloudy areas, the use of optical sensors could result in either missing or low-quality earth observation data depending on the cloud coverage [49]. Therefore, in this study, we tested the use of seasonal (i.e., dry, wet, and harvest) median and standard deviation backscatter of S1 data, together with two normalized difference-based backscatter indices (NDI), in improving the classification of mango orchards and other LULC classes (Table 3). This was based on the recommendation of

Schulz et al. [23] and Jin et al. [61], who also calculated these six S1-based variables for land use and crop mapping.

Table 3. Description of the vertical transmit/horizontal receive (VH) and vertical transmit/vertical receive (VV) backscatter bands and their indices ($n = 6$), derived from Sentinel-1 (S1) across the three seasonal clusters, i.e., wet, dry, and harvest seasons.

Band or Index	Wavelength or Formula
VV_p50	Vertically polarized backscatter with refined lee filter (seasonal mean)
VH_p50	Horizontally polarized backscatter with refined lee filter (seasonal mean)
VV_stdDev ³	The seasonal standard deviation of VV
VH_stdDev	The seasonal standard deviation of VH
NDI_VV	$(VV - VH)/(VV + VH)$
NDI_VH	$(VH - VV)/(VH + VV)$

³ stdDev = standard deviation.

The S1 mission supports data from a dual-polarization C-band SAR instrument at 5.405 GHz (C band). This compilation incorporates the S1 ground range detected (GRD) scenes, pre-processed by employing the S1 toolbox to create a calibrated and ortho-corrected product. We utilized S1 (S1A and S1B) GRD vertical transmit/horizontal receive (VH) and vertical transmit/vertical receive (VV) backscatters that were acquired during the ascending orbit (relative orbit: 1) using the interferometric wide swath (IW) for comparability of the backscatter intensity [61]. The GRD products were preferred because they detect the amplitude and are multi-looked to reduce the impact of speckles [62]. The S1 imagery (COPERNICUS/S1_GRD) was acquired already partially preprocessed at level 1 [62]. The stages in pre-processing involved generating level 1 data products that comprised the doppler centroid estimation, single look complex (SLC) focusing, and image post-processing to produce the SLC and GRD outputs, and mode-specific processing for the gathering of multiple sub-swath products [23,62]. The S1 toolbox in GEE was used to further preprocess the S1 backscatter data. This step involved speckle filtering, thermal noise removal, terrain correction, and radiometric calibration [50]. The terrain correction process was conducted using the shuttle radar topography mission (SRTM) of 30 m spatial resolution or the advanced spaceborne thermal emission and reflection (ASTER) digital elevation model (DEM) for areas greater than 60 degrees latitude where the SRTM is not available [62]. Additionally, a 3×3 filter window was applied to smoothen the backscatter data and reduce speckle effects using the Lee sigma filter [29,63].

3.5. Variable Selection and Importance

In total, we derived 81 S1- and S2-based predictor variables during the three seasons (i.e., dry, wet, and harvest) for mapping mango and other LULC classes in the study area. Specifically, we utilized 18 predictor variables from S1 and 63 from S2, which were subjected to a variable selection routine to reduce their dimensionality, multicollinearity, and redundancy [64]. Multicollinearity is often coupled with the restricted number of training samples (n) in comparison to the abundance of predictor variables (p) that often hinder (i.e., overfitting) the successful implementation of accurate predictive models when they are validated using an independent test dataset. As in our case, many predictor variables ($n = 81$) may have resulted in multicollinearity and information redundancy, which could negatively impact the classifier performance in mapping mango and other LULC classes. We generated three sets of predictor variables using three ML variable selection methods i.e., GRRF, VIF, and reliefF, and compared their performances. Initially, all the derived S1 and S2 variables ($n = 81$) were used for mapping the mango and other LULC classes before applying any variable selection. This is referred to as 'All' in the present study. The GRRF is one of the embedded methods that use decision trees, a similar concept to RF, but applies the importance scores generated from RF to guide the

variable selection process [40]. The importance value of a variable in RF is attained through the Gini index over all nodes, and across all the generated trees to facilitate the voting process of the variable selection [39]. Hence, GRRF provides a few variables that are most suitable for predicting target features of interest (e. g., LULC classes) from a dimension of variables [65]. The package “RRF” in R software (version 4.0.4, Vienna, Austria) [47] was used to perform the GRRF using a gamma value of 0.6 in the “CoefReg” [40,47]. On the other hand, VIF measures the inflation of the variance for the variables to determine multicollinearity among the variables. The VIF is directly calculated from a linear model with a focal numeric variable as a response using Equation (8):

$$\text{VIF}_i = \frac{1}{1 - R_i^2} \quad (8)$$

where i is the predictor variable and R^2 (R-squared) is the statistical measure that represents the proportion of the variance for the dependent variable that is explained by the independent variable.

The ‘vifcor’ function inherent in the ‘usdm’ package in R [47,66] was employed to perform variable selection using VIF. The ‘vifcor’ function picks variable pairs with high correlation and then removes the variable with the greatest VIF. The threshold was set at a Pearson correlation coefficient (r) of ≥ 0.7 . Fundamentally, a VIF value of more than 10 is a confirmation of collinearity among the variables [37]. The reliefF filter, as implemented in the R package ‘FSelector’ [67], was used as the third method for our variable selection experiment. The reliefF filter ranks the variables using the differences in the variables based on their nearest neighbors, thus the algorithm finds weights of continuous and discrete attributes according to a distance among all the variables [34].

In our variable selection experiment, we limited the number of the selected variables for each variable selection method to 18. This was informed by the maximum optimum number of variables that could be selected using the VIF method. Therefore, we restricted the other two methods to select the same number of variables (i.e., 18) as the VIF method for ease of comparison. Additionally, we ranked the importance (%) of the selected variables for mapping mango and other LULC classes using the inbuilt ‘variable importance’ procedure in the RF algorithm as provided in the GEE. For the ‘All’ variables, we selected the 18 most important predictor variables as ranked by the RF algorithm.

3.6. Mapping of Mango Orchards and Other Land Use and Land Cover (LULC) Classes

Two of the most widely used ML classifiers, i.e., SVM [41] and RF [39], were used to test the strength of combined S1- and S2-derived predictor variables for mapping mango and other LULC classes. Moreover, we opted to compare the performance of two efficient ML classifiers that have been previously tested for LULC mapping and have shown comparatively high accuracy [2,68]. These two ML classifiers are assumption-free methods that do not generally encounter prediction overfitting challenges. The SVM classifier divides the training data points by a hyperplane and maximizes the distance between a data point (e.g., a pixel) and the hyperplane to assign a class to such a data point. RF, on the other hand, uses decision trees that assign classes to data points following a predefined tree structure until the class labeling criteria are met. The algorithm uses a majority vote procedure that is derived from the decision trees to determine the final class label for each data point [39]. A total of 100 ‘trees’ were used to run the RF classification experiment after a parameter tuning process that demonstrated the classification accuracies would not considerably be improved when a higher number of ‘trees’ was used [23]. We used the default *mtry*, which is the square root of the number of predictor variables used. Similarly, we used the default SVM parameters in GEE, i.e., radial bases function (RBF) kernel type and a gamma value equivalent to one divided by the number of predictor variables (1/n features) for each run (1/81 for all features and 1/18 for the feature selected runs) [41]. We employed both SVM and RF classification experiments in the GEE [50].

3.7. Accuracy Assessment

Olofsson et al. [69] argued that the accuracy assessment results generated from the confusion classification matrix must not be the last step of the model evaluation but an essential step of the overall analysis of model accuracy. Therefore, the confusion classification matrix was used as a first step to calculate the interclass errors that were used to quantify the accuracy uncertainty using an area under class estimation method [69]. Specifically, we calculated overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA), and the Kappa coefficient to assess the performance of SVM and RF for mapping mango orchards and other coexisting LULC classes. Additionally, the area under each LULC class was used to build the confidence intervals at 95% to suggest the uncertainty of the area estimates of each variable selection method and classifier [69,70]. For a detailed description of the theoretic and mathematical constructs of the area under class accuracy assessment method, readers are referred to Olofsson et al. [69] and Card [70]. This method is widely accepted and has been used by several other studies [29,71,72]. A class-wise accuracy metric was subsequently calculated using F1-score, which leverages the importance of both the precision (UA) and recall (PA) in a single fused accuracy measure that ranges between 0 and 100% [73]. In other words, F1-score represents the harmonic mean between PA and UA for each class, as shown in Equation (9).

$$(F1)_i = (2 \times PA_i \times UA_i) \div (PA_i + UA_i) \quad (9)$$

Additionally, a McNemar's chi-square test was carried out to test for any statistically significant differences ($p \leq 0.05$) in the model performance among the different variable selection methods and the classifiers used [74].

4. Results

4.1. Variable Selections and Importance

The reliefF, GRRF, and VIF methods selected different S1- and S2-based predictor variables from the three seasons (Figure 3). However, most of the selected variables were from the wet season variables, followed by the dry and harvest seasons, respectively. In general, most of the selected variables were S1 and S2 vegetation indices and very few were wavebands. Additionally, there was no distinct consistency in the selected predictor variables between the four variable selection methods (Figure 3). The results also showed that most of the frequently selected predictor variables were derived from the S1 sensor (Table 4). The most frequently selected predictor variables (three out of four times) were the S1 VH backscatter band of the wet season (wet_VH_p50) and the VV backscatter band of the harvest season (harvest_VV_p50: Figure 3). Although the S1 derived bands and indices had the least representation in the total number of the selected predictor variables, they were the most frequently selected among the five most important predictor variables across the four variable selection methods (Figure 3).

Table 4. Frequency of Sentinel-1 and Sentinel-2 variables selected by the four variable selection methods used in this study.

Sensor	Once	Twice	Thrice	Four Times
Sentinel-1	13	7	2	0
Sentinel-2	21	9	0	0
Total	34	16	2	0

4.2. Mapping Mango and Other Land Use and Land Cover (LULC) Classes

Figure 4 shows the mapping results for the mango and other LULC classes in Zimbabwe using combinations of the four variable selection methods and two ML classification algorithms. Specifically, the RF and SVM classifiers mapped the distribution of mango differently, with the SVM underestimating the mango distribution in the northeastern side

of Mashonaland East; this was improved after combining SVM and VIF. In contrast, SVM mapped relatively more mango distribution in Mashonaland West as opposed to the RF when the relief and VIF variable selection methods were used (Figure 4), whereas the 'All' and GRRF variable selection methods showed relatively similar mango distribution trends in Mashonaland West when RF and SVM were employed. The results showed that water distribution was comparatively mapped out using all the combinations of the variable selection and classification methods with some confusion between water and forest in some parts of Mashonaland West using SVM. Moreover, different mapping patterns for the other LULC classes (e.g., grassland) were observed when different variable selection and classification methods were performed. Overall, the varying mango and other LULC mapping patterns indicate that each combination of variable selection and classification algorithms performed differently.

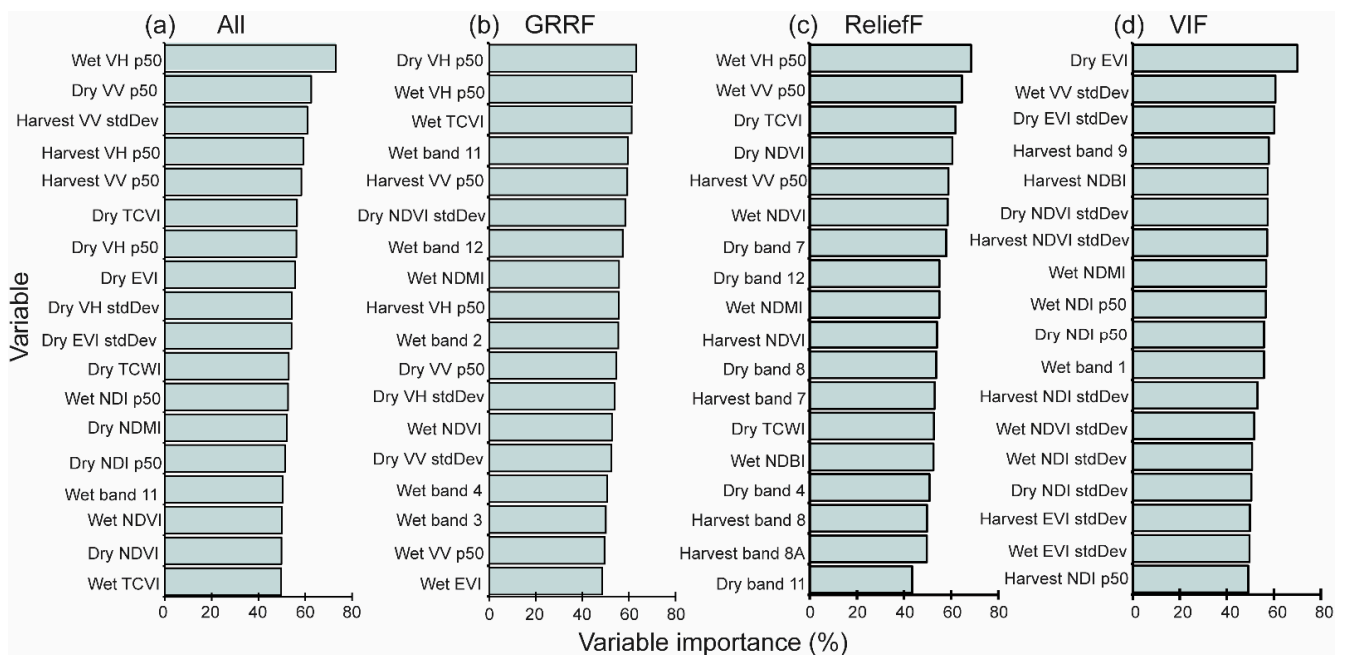


Figure 3. Variable selection and importance using four variable selection methods: (a) 'All', the 81 bands developed from Sentinel-1 and -2; (b) guided regularized random forest (GRRF); (c) relief factor (reliefF); and (d) variance inflation factor (VIF).

4.3. Analysis of the Classification Accuracies

The results revealed that the RF algorithm outperformed (test OA >75%, Kappa = 0.7) SVM (test OA = 74%, Kappa = 0.6) in mapping mango and other LULC classes in the four variable selection methods (Table 5). On the other hand, the best performing variable selection method as indicated by the test OA was the relief (test OA = 78.4%), while the least performing one was VIF (test OA = 66%). The OA achieved by the two classification algorithms was within the acceptable standard (~75%), and their strength and ability to improve the classification were demonstrated in the class-wise accuracies, as shown in Figure 5.

Although the summary of the overall accuracies of the different combinations of variable selection methods and the classification algorithms showed differences in accuracy levels, the McNemar test demonstrated that the differences were not statistically significant at $p \leq 0.05$ (Table 6).

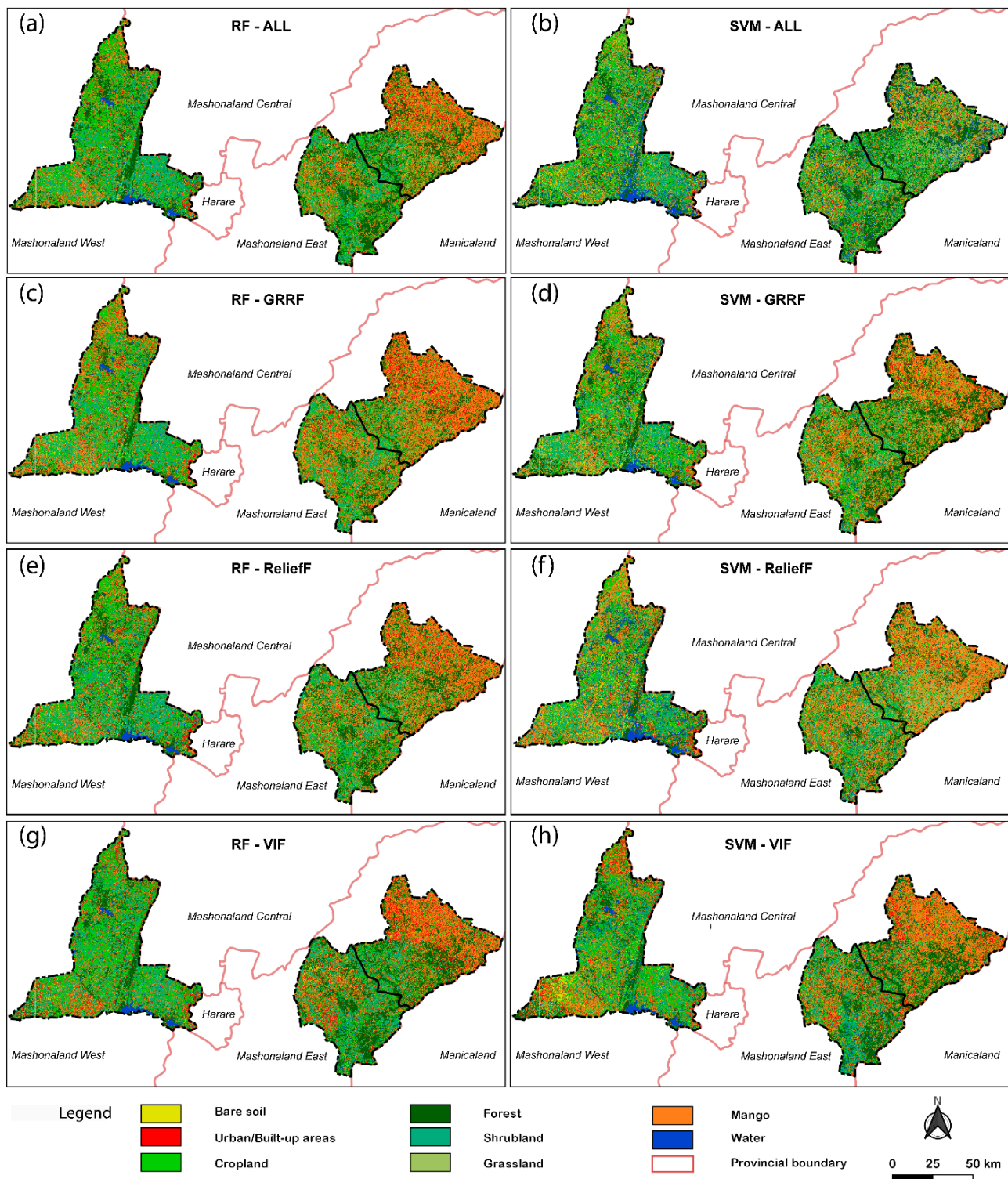


Figure 4. Comparative mango and other land use and land cover (LULC) maps for Mashonaland East and West in Zimbabwe that were produced using combinations of two classification algorithms and four variable selection methods: (a,b) random forest (RF) and support vector machines (SVM) with 'All' 81 bands (without variable selection), respectively; (c,d) RF and SVM with the guided regularized random forest (GRRF), respectively; (e,f) RF and SVM with reliefF, respectively; and (g,h) RF and SVM variance inflation factor (VIF), respectively.

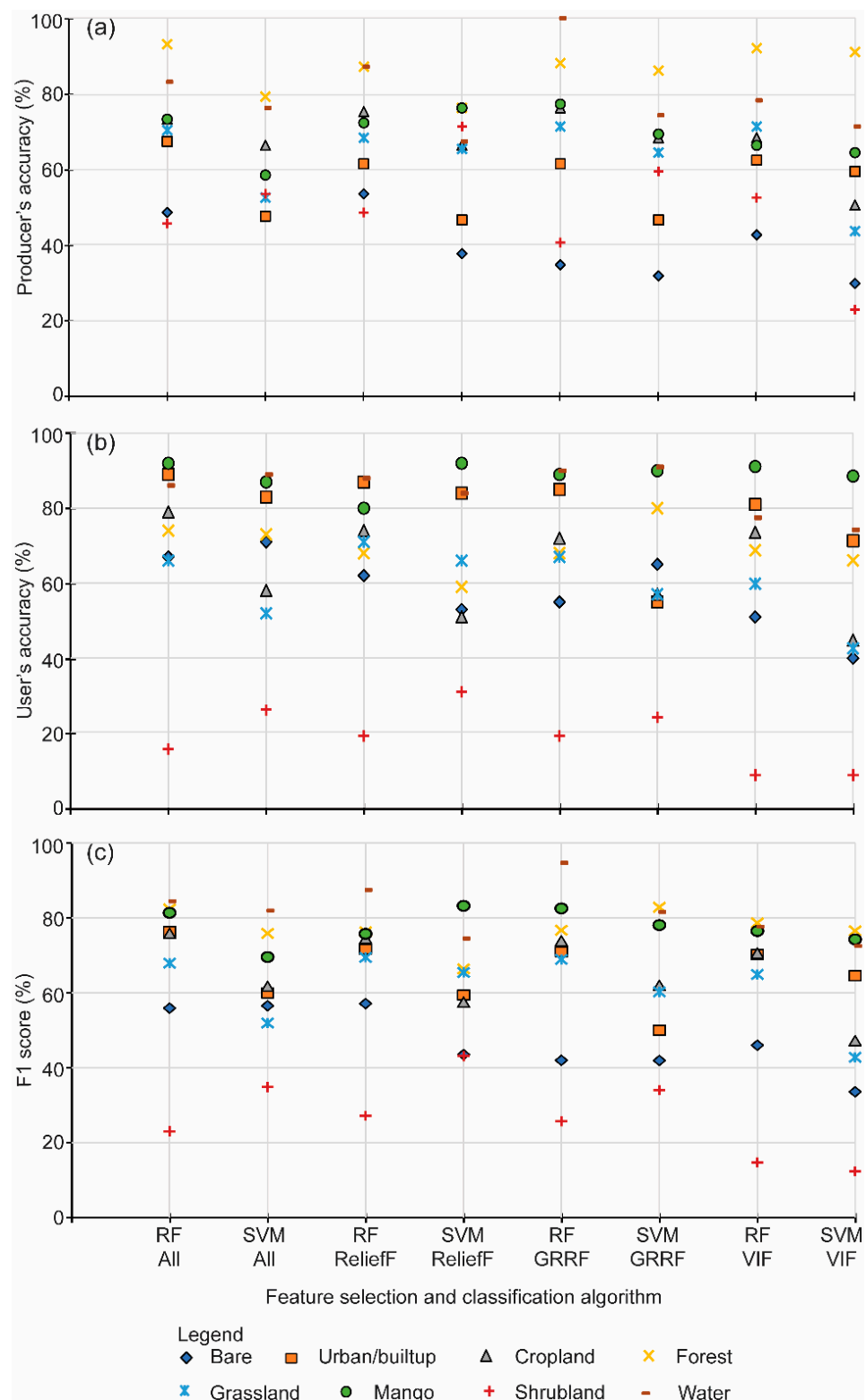


Figure 5. Comparative analysis of (a) producers’ accuracy, (b) users’ accuracy, and (c) F1 scores for the eight land use and land cover (LULC) classes using two machine learning (ML) classification algorithms (random forest (RF) and support vector machine (SVM)) and three variable selection methods, i.e., reliefF, guided regularized random forest (GRRF), and variance inflation factor (VIF), and using all the variables (‘ALL’). The classification accuracy metrics were calculated using the area under class method.

Table 5. Overall training and test accuracies for mapping mango and other land use and land cover (LULC) classes using the random forest (RF) and support vector machines (SVM) and three variable selection methods, i.e., reliefF, guided regularized random forest (GRRF), variance inflation factor (VIF), and using all variables ('ALL').

Classifier	RF			SVM		
Variable Selection Method	Training Overall Accuracy (%)	Test Overall Accuracy (%)	Kappa	Training Overall Accuracy (%)	Test Overall Accuracy (%)	Kappa
'All'	100	80	0.7	89	74	0.6
Relieff	100	78	0.7	76	74	0.6
GRRF	100	77	0.7	77	72	0.6
VIF	100	75	0.6	68	66	0.5

Table 6. Comparative analysis of the McNemar (χ^2) test ($p \leq 0.05$) for the eight possible classification combinations using two machine learning (ML) algorithms (random forest (RF) and support vector machine (SVM)) and three variable selection methods, i.e., reliefF, guided regularized random forest (GRRF), and variance inflation factor (VIF), and using all the variables ('ALL'). The values in the table show the p-values from the pairwise model comparisons.

	RF 'ALL'	SVM 'ALL'	RF ReliefF	SVM ReliefF	RF GRRF	SVM GRRF	RF VIF	SVM VIF
RF 'ALL'	– ⁴	–	–	–	–	–	–	–
SVM 'ALL'	0.6344	–	–	–	–	–	–	–
RF ReliefF	0.6738	0.6409	–	–	–	–	–	–
SVM ReliefF	0.6321	0.6008	0.6238	–	–	–	–	–
RF GRRF	0.6602	0.6279	0.6517	0.6265	–	–	–	–
SVM GRRF	0.6174	0.5867	0.6093	0.5854	0.6013	–	–	–
RF VIF	0.6436	0.6118	0.6352	0.6105	0.6269	0.6023	–	–
SVM VIF	0.5591	0.531	0.5517	0.5299	0.5444	0.5226	0.5357	–

⁴—means the value is not applicable or it is available in the other intersection dimension.

Figure 5 demonstrates the class-wise mapping accuracy metrics for the mango and other LULC classes for each variable selection method and classification algorithm. The PA, UA, and F1-scores were constantly above $80\% \pm 5\%$ for the mango and water classes, while the other LULC classes were within the $70\% \pm 5\%$ range, except for the shrubland class, which had relatively low accuracies (F1 scores less than 40%). In particular, the RF and SVM models and the three variable selection methods successfully classified the mango, water, and urban/built-up classes with very high accuracy (F1 score above 75%), except when the VIF was used in combination with SVM (Figure 5c). Additionally, improvements were observed in the F1 scores of these three classes while using the reliefF and GRRF with the RF model ($85\% \pm 4\%$). Results revealed that an F1 score accuracy range of 70–80% for all the variable selection methods and classification algorithms was achieved for the mango class (Figure 5c). Although the shrubland class had relatively low accuracies across the different variable selection and classification methods, it was observed that using reliefF and SVM resulted in the relatively better mapping performance of shrublands as compared to the other methods (Figure 5). It was also interesting to note that all the variable selection and classification algorithms mapped the urban/built-up class with high accuracies ($\pm 75\%$). It was thus observed that each classifier and variable selection method had its advantages and limitations in mapping certain classes of LULC.

4.4. Unbiased Area Estimation

The unbiased area estimate calculated from the best performing combination of the reliefF and RF classification scenario revealed that mango orchards cover approximately 18% ($292,232 \pm 29,358$ ha) of the total area coverage in the three districts i.e., Zvimba, Murehwa, and Mutoko. The largest area was demonstrated to be covered by croplands (33%), while the least coverage was by water, bare areas, and shrubland classes (Table 7).

Table 7. The area estimates and approximate 95% confidence intervals (CIs) of the mango and other land use and land cover (LULC) classes using the area under class method of the highest obtained accuracy, i.e., the reliefF variable selection method and the random forest (RF) classification algorithm.

Class	Unbiased Area Estimate (ha)	Area (%)	95% CI of the Unbiased Area Estimate (ha)
Mango	292,232	18	$\pm 29,358$
Built-up	151,147	9	$\pm 20,141$
Cropland	529,021	33	$\pm 37,160$
Forest	271,183	17	$\pm 34,127$
Grassland	256,377	16	$\pm 31,994$
Bare soil	26,199	2	± 9506
Shrubland	54,771	3	$\pm 20,841$
Water	28,797	2	± 5606
Total	1,609,726	100	

5. Discussion

We tested the strength of S1 backscatter, S2 reflectance bands, and their indices and derivatives for distinguishing mango orchard and the other seven LULC classes in a semi-arid environment in southern Africa using four variable selection and two ML classification methods. Our mango mapping approach builds on the recommendations of earlier studies that targeted large-scale, crop-specific mapping using the freely available Sentinel dataset [49]. Previous attempts for mapping mango trees resulted in high misclassification (i.e., low accuracy) that could be due to the weakness of the data and classification methods used [1]. Thus, in the present study, we advanced the geospatial methodologies for mapping mango as one of the important tree crops for food and nutrition security and income generation in Africa and produced plot-level (i.e., orchard) thematic maps in a heterogeneous landscape. Furthermore, the advancements in data analytic tools, such as the GEE approach used in this study, are reproducible and semiautomatic. Thus, the mapping approach used provides unique opportunities to improve the detection and monitoring of mango orchards of variable sizes, which might not be feasible using the conventional remote sensing mapping tools [30,53]. The GEE offers advantages such as ML and parallel processing of large satellite-based datasets such as those used in this study ($n = 81$), memory efficiency, and fast image processing power [50].

The results demonstrated the usefulness of combining S1 SAR and S2 optical variables with ML variable selection and classification methods for mapping mango and other LULC classes over a wide area in a semi-arid environment. Additionally, the results showed valuable marks for establishing relevant and highly important predictor variables that are necessary for mapping fruit tree crops, and LULC in general, in such complex and heterogeneous semi-arid environments. The aggregation of distinct Sentinel image dates to capture the yearly inter-seasonal spectral variations among the studied LULC classes provided a huge opportunity to improve our mapping accuracies [23,49]. This is in confirmatory of other studies that demonstrated the utility of the Sentinels' spatial (10 m) and temporal (6 days for S1 and 5 days for S2) resolutions, as well as relevant spectral and backscatter variables in distinguishing different vegetation classes (e.g., fruit tree, cropland, and forest trees) and non-vegetation (e.g., bare land and water) classes albeit the complexity and heterogeneity of semi-arid environments [30]. Furthermore, the mango

detection approach was sound and innovative as it leveraged on initially generating many ($n = 81$) integrated S1 and S2 predictor variables across three seasons (i.e., dry, wet, and harvest), then applying robust variable selection methods to select a few, yet relevant ones to map mango and other LULC classes using two ML classifiers (RF and SVM).

This is the first attempt aimed at mapping mango orchards at relatively large geographic scales in Africa. However, the use of combined S1 and S2 datasets for mapping LULC classes including tree crops (e.g., avocado and coffee) is well documented [29,49]. Although the incorporation of S1 and S2 datasets in mapping LULC is not new, it could be regarded as innovative because it was developed and advanced to suit different crops (e.g., mango), regions, and agro-ecologies [23,49]. Therefore, our study contributes to this research portfolio by mapping mango orchards. Furthermore, most of the earlier studies on LULC mapping have exclusively relied on the use of optical satellite data [22,75]. Unlike the present study, studies that have used a combination of optical and SAR Sentinel data for LULC mapping were carried out on relatively smaller geographic footprints [29]. The combination of S1 and S2 data showed that the mango and other LULC classes were mapped with high OA ($75 \pm 5\%$). Likewise, our mango orchard class was detected with an F1 score of $>70\%$. This is following the findings of earlier studies that integrated S1 and S2 variables for mapping crop and cropping patterns at a landscape scale and obtained classification accuracies ranging between 70% and 90% [23,29,76]. This could be attributed to the added advantage of S1, i.e., the non-reliance on cloud-free days and the smaller wavelength of the C-band (3–5 cm), allowing for a temporally continual data source to augment the optical data, particularly in the cloudy tropics and sub-tropics. Additionally, the area under class analysis showed that $\sim 18\%$ of the study area was covered by mango. This result highlighted the huge contribution and spatial coverage that mango has within the communities in Zvimba, Murehwa, and Mutoko. This concurred with earlier studies that have established that mango trees have been naturalized in Zimbabwe and can be found both in the wild and in domesticated systems [77]. The natural forest class had a larger coverage compared to the mango class, thus we assume that some forest trees perhaps mixed with naturalized mango trees could have been classified as mango, as shown by the large range of the confidence interval, i.e., $\pm 29,358$ ha.

The results also showed low PA and UA for mapping the shrubland class. This could have contributed to lowering the OA obtained in this study. The low grassland mapping accuracies obtained in this study could be attributed to the relatively low number of ground truth points for the grassland and the expected spectral similarity between grassland and bare soil [78]. Probably an addition of training samples for this class would have reduced the mapping errors by reducing the within-class variability [79,80]. Additionally, the classification approach used in the present study was able to capture the spectral differences among the eight classes within a highly heterogeneous landscape. For instance, neighboring mango orchards are often patchworked along with annual (e.g., maize) and perennial (e.g., guava) crops. Some other mango orchards were within a village setup in the present study area, hence they are surrounded by built-up and crop classes. Despite these intricacies, both ML classifiers used in this study produced a good agreement between the test and predicted LULC observations as indicated by the Kappa metric.

It is interesting to note that the present study presents an opportunity to comparatively assess the performance of RF and SVM classifiers for mapping mango, as a tree crop, and other LULC classes in complex and heterogeneous agroecology. An in-depth analysis showed that RF consistently outperformed SVM as assessed using the OA and class-wise accuracy metrics, except for the grassland class where SVM achieved a relatively better F1 score. This agrees with several studies that reported a better performance of RF for mapping LULC classes as opposed to other ML classifiers such as SVM [57]. However, other studies have found that SVM outperformed RF in mapping LULC classification using S2 data and other ancillary predictor variable types [81]. These disparities in the performance of ML classifiers are expected and are predominantly affected by the selected predictor variables and the number and distribution of ground truth observations and the parameters and

settings used for executing the classifier. Therefore, what can be derived from a critical comparison among these studies is that the differences in their performance in most cases are not statistically substantial, as evidenced by the McNemar comparative analysis.

The S1 and S2 imagery are rich with information that is optimum and suitable to predict LULC classes. Specifically, the variable selection and importance experiment presented strong evidence of the usefulness and relevance inherent in the S1-based variables for mapping mango and other LULC classes. This is indicated by the number of S1-based variables that dominated the top-ranked variables for mapping mango and other LULC classes compared to the S2-derived variables. The importance and dominance of the S1 variables could be attributed to SAR's independence to the time of the day or the seasonal changes compared to the optical sensors such as S2. However, by adding the temporal dimension in the S1 and S2 data via aggregating the seasonal variability instead of using single imaging dates, we were able to eliminate the inherent bias toward selecting S1-based variables [23]. Hence, S2-based variables were also well represented among the 18 selected predictor variables for classifying LULC. Likewise, Jin et al. [61] demonstrated the importance of both S1- and S2-based variables for mapping cropland presence, maize presence, and maize yields for the main 2017 maize season in Kenya and Tanzania. The variable selection and importance approach reduced the dimensionality of the initially generated variables ($n = 81$) by 78%. Studies have shown that using a few relevant predictor variables in a classification experiment reduces the overfitting and computational time required for executing the classification algorithm itself [64]. Additionally, this could permit a better transferability or extrapolation of our mapping method in similar agro-ecologies. In particular, the approach developed in this study could be used to detect and map mango orchards in other areas in Zimbabwe and elsewhere. Notwithstanding, the transferability of the approach to other points in time and space was not tested. Furthermore, for other areas with similar climatic patterns and agro-ecologies, it is speculated that the selected predictor variables would perform similarly to their performance for detecting mango and other LULC classes in the present study. Thus, we recommend that studies that aim to map mango and other tree crops in tropical and semi-arid agro-ecologies should integrate both SAR and optical seasonal composite datasets during dry, wet, and harvest seasons that are defined in the localized target climates. Specifically, the identification and definition of the seasons during which SAR and optical data should be captured to map mango and other LULC classes should solely depend on the number of rainy seasons, onset, cessation of the precipitation, and its distribution in the specific areas [82].

Overall, the mango and other LULC mapping results could serve as a baseline for future LULC prediction studies including change detection. Likewise, the developed LULC maps could be used together with other environmental and vegetation information as input predictor variables for assessing the habitat suitability of mango pests and pollinators using species distribution models. Moreover, the LULC maps could also be advisory tools for extension services related to, for instance, crop inventory, husbandry, postharvest interventions, insurance, and marketing. The present study did not particularly differentiate between mango age groups, hence future studies should look at the possibility of using integrated multi-date SAR and optical datasets for discriminating distinct mango age groups or estimation of individual tree numbers. This would provide better and more accurate information as compared to a general LULC map for interventions such as harvest planning, marketing, and postharvest. On the other hand, the LULC mapping results could be baseline inputs in studies that look at land degradation and fragmentation, such as deforestation expansion of agriculture. These LULC pathways would be exceptionally useful in building a timeline of a spatial continuous proxy for, e.g., agricultural expansion and forest vulnerability.

6. Conclusions

Our study demonstrated and provided an innovative mango and other LULC mappings approach for heterogeneous landscapes in semi-arid regions. Overall, the findings

provide baseline information appropriate to perform LULC change analysis as well as a predictor variable input in species distribution models such as in insect pests and pollinators. Detecting the mango class benefitted from the strength of cloud computing using multiple indices together with seasonal temporal clustering of the remotely sensed data. However, the results of the present study suggested that using more data did not always generate better results, hence the need for variable selection to reduce overfitting and the time needed for the analysis. The study also showed that there were shortcomings in the classification of the shrubland class, probably due to a low number of reference points and spectral similarity with the bare soil and grassland classes. Although the focus of the study was on mango orchards, most reference points belonged to one class, overestimating this class compared to all of the other seven classes. Future studies could investigate combining such similar classes to improve the classification accuracies. Although RF proved superior to SVM in mango detection, their differences were not significant, hence SVM should also be considered during LULC classification using robust variable selection methods such as reliefF. Together with other earlier studies, the performance of the classifier and variable selection approach was not uniform, thus yielding different classification outcomes. Therefore, we suggest the use of an ensemble approach to harness the strengths of each classifier and each set of predictor variables.

Author Contributions: Conceptualization, B.T.M. and E.M.A.-R.; methodology, B.T.M.; software, B.T.M.; validation, E.M.A.-R., H.E.Z.T., S.N., L.D.M.M., C.C.N. and S.A.M.; formal analysis, B.T.M.; investigation, B.T.M., E.M.A.-R., H.E.Z.T. and S.N.; resources, S.N. and S.A.M.; data curation, B.T.M.; writing—original draft preparation, B.T.M.; writing—review and editing, E.M.A.-R., H.E.Z.T., S.N., L.D.M.M., C.C.N. and S.A.M.; visualization, E.M.A.-R. and H.E.Z.T.; supervision, S.A.M.; project administration, S.N.; funding acquisition, S.A.M., All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the financial support for this research by the following organizations and agencies: International Development Research Centre-Centre de recherches pour le développement international (IDRC-CRDI) and the Australian Centre for International Agricultural Research (ACIAR) (grant number 109040); the Swedish International Development Cooperation Agency (Sida); the Swiss Agency for Development and Cooperation (SDC); the Federal Democratic Republic of Ethiopia; and the Government of the Republic of Kenya. “The views expressed herein do not necessarily reflect the official opinion of the donors.”

Data Availability Statement: Data used in this study is available from the corresponding author on request.

Acknowledgments: Sincere gratitude is also extended to the extension officers in Zimbabwe for their support in data collection. The authors also acknowledge Benard Malenge for producing the maps.

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

References

1. Luo, H.X.; Dai, S.P.; Li, M.F.; Liu, E.P.; Zheng, Q.; Hu, Y.Y.; Yi, X.P. Comparison of machine learning algorithms for mapping mango plantations based on Gaofen-1 imagery. *J. Integr. Agric.* **2020**, *19*, 2815–2828. [[CrossRef](#)]
2. Forkuor, G.; Dimobe, K.; Serme, I.; Tondoh, J.E. Landsat-8 vs. Sentinel-2: Examining the added value of sentinel-2’s red-edge bands to land-use and land-cover mapping in Burkina Faso. *GISci. Remote Sens.* **2018**, *55*, 331–354. [[CrossRef](#)]
3. Dobrota, C.T.; Carpa, R.; Butiuc-Keul, A. Analysis of designs used in monitoring crop growth based on remote sensing methods. *Turkish J. Agric. For.* **2021**, *45*, 730–742. [[CrossRef](#)]
4. Alkan, A.; Abdullah, M.Ü.; Abdullah, H.O.; Assaf, M.; Zhou, H. A smart agricultural application: Automated detection of diseases in vine leaves using hybrid deep learning. *Turkish J. Agric. For.* **2021**, *45*, 717–729. [[CrossRef](#)]
5. Zingore, K.M.; Sithole, G.; Abdel-Rahman, E.M.; Mohamed, S.A.; Ekesi, S.; Tanga, C.M.; Mahmoud, M.E.E. Global risk of invasion by *Bactrocera zonata*: Implications on horticultural crop production under changing climatic conditions. *PLoS ONE* **2020**, *15*, e0243047. [[CrossRef](#)]
6. FAO; IFAD; UNICEF; WFP; WHO. The State of Food Security and Nutrition in the World 2021. In *Transforming Food Systems for Food Security, Improved Nutrition and Affordable Healthy Diets for All*; FAO: Rome, Italy, 2021; ISBN 9789251329016.
7. FAO. *How to Feed the World in 2050. Insights from an Expert Meet*; FAO: Rome, Italy, 2009; pp. 1–35. [[CrossRef](#)]
8. FAO. *Fruit and Vegetables—Your Dietary Essentials*; FAO: Rome, Italy, 2020; ISBN 9789251337097.

9. Durán Zuazo, V.H.; Rodríguez Pleguezuelo, C.R.; Gálvez Ruiz, B.; Gutiérrez Gordillo, S.; García-Tejero, I.F. Water use and fruit yield of mango (*Mangifera indica* L.) grown in a subtropical Mediterranean climate. *Int. J. Fruit Sci.* **2019**, *19*, 136–150. [CrossRef]
10. FAOSTAT Crops and Livestock Products: Mangoes, Mangosteens and Guavas. Available online: <https://www.fao.org/faostat/en/#data/QCL> (accessed on 3 February 2022).
11. Mujuka, E.; Mburu, J.; Ogotu, A.; Ambuko, J.; Magambo, G. Consumer awareness and willingness to pay for naturally preserved solar-dried mangoes: Evidence from Nairobi, Kenya. *J. Agric. Food Res.* **2021**, *5*, 100188. [CrossRef]
12. Mithöfer, D. Economics of Indigenous Fruit Tree Crops in Zimbabwe. Ph.D. thesis, University of Hannover, Hannover, Germany, 2004.
13. Ekesi, S.; Mohamed, S.A.; Meyer, M. *Fruit Fly Research and Development in Africa—Towards A Sustainable Management Strategy to Improve Horticulture*; Springer: Cham, Switzerland, 2016; ISBN 9783319432243.
14. Ekesi, S.; Billah, M.K.; Nderitu, P.W.; Lux, S.A.; Rwomushana, I. Evidence for competitive displacement of *Ceratitis cosyra* by the invasive fruit fly *Bactrocera invadens* (Diptera: Tephritidae) on mango and mechanisms contributing to the displacement. *J. Econ. Entomol.* **2009**, *102*, 981–991. [CrossRef]
15. Nankinga, C.M.; Isabirye, B.E.; Muyinza, H.; Rwomushana, I.; Stevenson, P.C.; Mayamba, A. Fruit fly infestation in mango: A threat to the Horticultural sector in Uganda. *Uganda J. Agric. Sci.* **2014**, *15*, 1–14.
16. Dobrini, D.; Gašparovi, M.; Medak, D. Sentinel-1 and 2 Time-Series for vegetation mapping using random forest classification: A Case Study of Northern Croatia. *Remote Sens.* **2021**, *13*, 2321. [CrossRef]
17. Campbell, J.; Wynne, R. *Introduction to Remote Sensing*, 5th ed.; Guilford Press: New York, NY, USA, 2007; Volume 136, ISBN 9781609181765.
18. Immitzer, M.; Vuolo, F.; Atzberger, C. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sens.* **2016**, *8*, 166. [CrossRef]
19. Ramoelo, A.; Cho, M.; Mathieu, R.; Skidmore, A.K. Potential of Sentinel-2 spectral configuration to assess rangeland quality. *J. Appl. Remote Sens.* **2015**, *9*, 094096. [CrossRef]
20. Chemura, A.; Mutanga, O.; Odindi, J.; Kutuywayo, D. Mapping spatial variability of foliar nitrogen in coffee (*Coffea arabica* L.) plantations with multispectral Sentinel-2 MSI data. *ISPRS J. Photogramm. Remote Sens.* **2018**, *138*, 1–11. [CrossRef]
21. Belgiu, M.; Csillik, O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* **2018**, *204*, 509–523. [CrossRef]
22. Mudereri, B.T.; Chitata, T.; Mukanga, C.; Mupfiga, T.; Gwatirisa, C.; Dube, T. Can biophysical parameters derived from Sentinel-2 space-borne sensor improve land cover characterisation in semi-arid regions? *Geocarto Int.* **2021**, *36*, 2204–2223. [CrossRef]
23. Schulz, D.; Yin, H.; Tischbein, B.; Verleysdonk, S.; Adamou, R.; Kumar, N. Land use mapping using Sentinel-1 and Sentinel-2 time series in a heterogeneous landscape in Niger, Sahel. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 97–111. [CrossRef]
24. ESA Sentinel-2 Products Specification. Available online: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/data-products> (accessed on 30 April 2019).
25. Cheng, K.; Wang, J. Forest-type classification using time-weighted dynamic timewarping analysis in mountain areas: A case study in southern China. *Forests* **2019**, *10*, 1040. [CrossRef]
26. Shoko, C.; Mutanga, O. Examining the strength of the newly-launched Sentinel 2 MSI sensor in detecting and discriminating subtle differences between C3 and C4 grass species. *ISPRS J. Photogramm. Remote Sens.* **2017**, *129*, 32–40. [CrossRef]
27. West, H.; Quinn, N.; Horswell, M. Remote sensing for drought monitoring & impact assessment: Progress, past challenges and future opportunities. *Remote Sens. Environ.* **2019**, *232*, 111291. [CrossRef]
28. Pandey, A.C.; Kaushik, K.; Parida, B.R. Google Earth Engine for large-scale flood mapping using SAR data and impact assessment on agriculture and population of Ganga-Brahmaputra basin. *Sustainability* **2022**, *14*, 4210. [CrossRef]
29. Aduvukha, G.R.; Abdel-rahman, E.M.; Sichangi, A.W.; Makokha, G.O.; Landmann, T.; Mudereri, B.T.; Tonnang, H.E.Z.; Dubois, T. Cropping pattern mapping in an agro-natural heterogeneous landscape using Sentinel-2 and Sentinel-1 satellite datasets. *Agriculture* **2021**, *11*, 530. [CrossRef]
30. Gxokwe, S.; Dube, T.; Mazvimavi, D. Leveraging Google Earth Engine platform to characterize and map small seasonal wetlands in the semi-arid environments of South Africa. *Sci. Total Environ.* **2021**, *803*, 150139. [CrossRef] [PubMed]
31. Nhu, V.H.; Mohammadi, A.; Shahabi, H.; Ahmad, B.B.; Al-Ansari, N.; Shirzadi, A.; Geertsema, M.; Kress, V.R.; Karimzadeh, S.; Kamran, K.V.; et al. Landslide detection and susceptibility modeling on cameron highlands (Malaysia): A comparison between random forest, logistic regression and logistic model tree algorithms. *Forests* **2020**, *11*, 830. [CrossRef]
32. Ahmad, A.; Gilani, H.; Ahmad, S.R. Forest aboveground biomass estimation and mapping through high-resolution optical satellite imagery—A literature review. *Forests* **2021**, *12*, 914. [CrossRef]
33. Georganos, S.; Grippa, T.; Vanhuysse, S.; Lennert, M.; Shimoni, M.; Kalogirou, S.; Wolff, E. Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. *GIScience Remote Sens.* **2018**, *55*, 221–242. [CrossRef]
34. Urbanowicz, R.J.; Meeker, M.; La Cava, W.; Olson, R.S.; Moore, J.H. Relief-based feature selection: Introduction and review. *J. Biomed. Inform.* **2018**, *85*, 189–203. [CrossRef]
35. Urszula, S.; Zielosko, B.; Jain, L.C. *Advances in Feature Selection for Data and Pattern Recognition*; Springer: Cham, Switzerland, 2018; Volume 138, ISBN 9783319136592.
36. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [CrossRef]

37. Mtengwana, B.; Dube, T.; Mudereri, B.T.; Shoko, C. Modeling the geographic spread and proliferation of invasive alien plants (IAPs) into new ecosystems using multi-source data and multiple predictive models in the Heuningnes catchment, South Africa. *GIScience Remote Sens.* **2021**, *58*, 483–500. [[CrossRef](#)]
38. De Leeuw, J. Journal of statistical software. *Wiley Interdiscip. Rev. Comput. Stat.* **2009**, *1*, 128–129. [[CrossRef](#)]
39. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
40. Deng, H.; Runger, G. Gene selection with guided regularized random forest. *Pattern Recognit.* **2013**, *46*, 3483–3489. [[CrossRef](#)]
41. Vapnik, V. Estimation of Dependences Based on Empirical Data. *Nauk. Moscow Transl.* **1979**, *27*, 5165–5184.
42. Muthoni, F.K.; Odongo, V.O.; Ochieng, J.; Mugalavai, E.M.; Mourice, S.K.; Hoesche-Zeledon, I.; Mwila, M.; Bekunda, M. Long-term spatial-temporal trends and variability of rainfall over Eastern and Southern Africa. *Theor. Appl. Climatol.* **2019**, *137*, 1869–1882. [[CrossRef](#)]
43. Mugandani, R.; Wuta, M.; Makarau, A.; Chipindu, B. RE-Classification of Agro-ecological regions of Zimbabwe in conformity with climate variability and change. *Afr. Crop Sci. J.* **2012**, *20*, 361–369.
44. Kuri, F.; Murwira, A.; Murwira, K.S.; Masocha, M. Accounting for phenology in maize yield prediction using remotely sensed dry dekads. *Geocarto Int.* **2018**, *33*, 723–736. [[CrossRef](#)]
45. Ouma, T.; Kavoo, A.; Wainaina, C.; Ogunya, B.; Karanja, M.; Kumar, P.L.; Shah, T. Open data kit (ODK) in crop farming: Mobile data collection for seed yam tracking in Ibadan, Nigeria. *J. Crop Improv.* **2019**, *33*, 605–619. [[CrossRef](#)]
46. Tonnang, H.E.Z.; Balemi, T.; Masuki, K.F.; Mohammed, I.; Adewopo, J.; Adnan, A.A.; Mudereri, B.T.; Vanlauwe, B.; Craufurd, P. Rapid acquisition, management, and analysis of spatial Maize (*Zea mays* L.) phenological data—Towards ‘Big Data’ for agronomy transformation in Africa. *Agronomy* **2020**, *10*, 1363. [[CrossRef](#)]
47. R Core. *Team R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria; Available online: <https://www.R-project.org/> (accessed on 18 November 2021).
48. Kuhn, M. Package ‘caret’ R topics documented: CRAN Repos. 2020. Available online: <https://CRAN.R-project.org/package=caret> (accessed on 18 November 2021).
49. 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](#)]
50. 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](#)]
51. WFP. *Seasonal Overview and Regional Southern African Vulnerability Analysis (2020/2021)*; WFP: Johannesburg, South Africa, 2021.
52. 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](#)]
53. Mudereri, B.T.; Abdel-Rahman, E.M.; Dube, T.; Niassy, S.; Khan, Z.; Tonnang, H.E.Z.; Landmann, T. A two-step approach for detecting Striga in a complex agroecological system using Sentinel-2 data. *Sci. Total Environ.* **2021**, *762*, 143151. [[CrossRef](#)] [[PubMed](#)]
54. Henrich, V.; Krauss, G.; Gotze, C.; Sandow, C.; IDB. Entwicklung einer Datenbank für Fernerkundungsindizes. Available online: www.indexdatabase.de (accessed on 18 November 2021).
55. Xue, J.; Su, B. Significant remote sensing vegetation indices: A review of developments and applications. *J. Sens.* **2017**, *2017*, 1353691. [[CrossRef](#)]
56. Chemura, A.; Mutanga, O.; Dube, T. Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. *Precis. Agric.* **2017**, *18*, 859–881. [[CrossRef](#)]
57. Camargo, F.F.; Sano, E.E.; Almeida, C.M.; Mura, J.C.; Almeida, T. A comparative assessment of machine-learning techniques for land use and land cover classification of the Brazilian tropical savanna using ALOS-2/PALSAR-2 polarimetric images. *Remote Sens.* **2019**, *11*, 1600. [[CrossRef](#)]
58. Odipo, V.O.; Nickless, A.; Berger, C.; Baade, J.; Urbazaev, M.; Walther, C.; Schmullius, C. Assessment of aboveground woody biomass dynamics using terrestrial laser scanner and L-band ALOS PALSAR data in South African Savanna. *Forests* **2016**, *7*, 294. [[CrossRef](#)]
59. Brown, C.; Daniels, A.; Boyd, D.S.; Sowter, A.; Foody, G.; Kara, S. Investigating the potential of radar interferometry for monitoring rural artisanal cobalt mines in the democratic republic of the congo. *Sustainability* **2020**, *12*, 9834. [[CrossRef](#)]
60. Patel, P.; Srivastava, H.S.; Panigrahy, S.; Parihar, J.S. Comparative evaluation of the sensitivity of multi-polarized multi-frequency SAR backscatter to plant density. *Int. J. Remote Sens.* **2006**, *27*, 293–305. [[CrossRef](#)]
61. Jin, Z.; Azzari, G.; You, C.; Di Tommaso, S.; Aston, S.; Burke, M.; Lobell, D.B. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sens. Environ.* **2019**, *228*, 115–128. [[CrossRef](#)]
62. ESA Sentinel Online: User Guides Sentinel-1 SAR Product Types and Processing Levels. Available online: <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/product-types-processing-levels/level-1> (accessed on 15 November 2021).
63. Lee, J.S.; Wen, J.H.; Ainsworth, T.L.; Chen, K.S.; Chen, A.J. Improved sigma filter for speckle filtering of SAR imagery. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 202–213. [[CrossRef](#)]
64. Izquierdo-Verdiguier, E.; Zurita-Milla, R. An evaluation of guided regularized random forest for classification and regression tasks in remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *88*, 102051. [[CrossRef](#)]

65. Mureriwa, N.; Adam, E.; Sahu, A.; Tesfamichael, S. Examining the spectral separability of *Prosopis glandulosa* from co-existent species using field spectral measurement and guided regularized random forest. *Remote Sens.* **2016**, *8*, 144. [CrossRef]
66. Naimi, B.; Hamm, N.A.S.; Groen, T.A.; Skidmore, A.K.; Toxopeus, A.G. Where is positional uncertainty a problem for species distribution modelling? *Ecography* **2014**, *37*, 191–203. [CrossRef]
67. Romanski, P.; Kotthoff, L.; Schratz, P. FSelector: Selecting Attributes. R package version 0.33. 2021. Available online: <https://CRAN.R-project.org/package=FSelector> (accessed on 18 November 2021).
68. Mushore, T.D.; Mutanga, O.; Odindi, J.; Dube, T. Linking major shifts in land surface temperatures to long term land use and land cover changes: A case of Harare, Zimbabwe. *Urban Clim.* **2017**, *20*, 120–134. [CrossRef]
69. Olofsson, P.; Foody, G.M.; Stehman, S.V.; Woodcock, C.E. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sens. Environ.* **2013**, *129*, 122–131. [CrossRef]
70. Card, D.H. Using known map category marginal frequencies to improve estimates of thematic map accuracy. *Photogramm. Eng. Remote Sens.* **1982**, *48*, 431–439.
71. Stehman, S.V.; Mousoupetros, J.; McRoberts, R.E.; Næsset, E.; Pengra, B.W.; Xing, D.; Horton, J.A. Incorporating interpreter variability into estimation of the total variance of land cover area estimates under simple random sampling. *Remote Sens. Environ.* **2022**, *269*, 112806. [CrossRef]
72. Shibia, M.G.; Röder, A.; Fava, F.P.; Stellmes, M.; Hill, J. Integrating satellite images and topographic data for mapping seasonal grazing management units in pastoral landscapes of eastern Africa. *J. Arid Environ.* **2022**, *197*. [CrossRef]
73. Mudereri, B.T.; Mukanga, C.; Mupfiga, E.T.; Gwatirisa, C.; Kimathi, E.; Chitata, T. Analysis of potentially suitable habitat within migration connections of an intra-African migrant-the Blue Swallow (*Hirundo atrocaerulea*). *Ecol. Inform.* **2020**, *57*, 101082. [CrossRef]
74. McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **1947**, *12*, 153–157. [CrossRef]
75. Zvobgo, L.; Tsoka, J. Deforestation rate and causes in Upper Manyame Sub-Catchment, Zimbabwe: Implications on achieving national climate change mitigation targets. *Trees For. People* **2021**, *5*, 100090. [CrossRef]
76. Csillik, O.; Belgiu, M. Cropland mapping from Sentinel-2 time series data using object-based image analysis. In Proceedings of the 20th AGILE International Conference on Geographic Information Science Societal Geo-Innovation Celebrating, Wageningen, The Netherlands, 9–12 May 2017.
77. CABI Mango: *Mangifera Indica*. Available online: <https://www.cabi.org/isc/datasheet/34505> (accessed on 7 February 2022).
78. Holden, P.B.; Rebelo, A.J.; New, M.G. Mapping invasive alien trees in water towers: A combined approach using satellite data fusion, drone technology and expert engagement. *Remote Sens. Appl. Soc. Environ.* **2021**, *21*, 100448. [CrossRef]
79. Royimani, L.; Mutanga, O.; Odindi, J.; Zolo, K.S.; Sibanda, M.; Dube, T. Distribution of *Parthenium hysterophoru* L. with variation in rainfall using multi-year SPOT data and random forest classification. *Remote Sens. Appl. Soc. Environ.* **2019**, *13*, 215–223. [CrossRef]
80. Makaya, N.P.; Mutanga, O.; Kiala, Z.; Dube, T.; Seutloali, K.E. Assessing the potential of Sentinel-2 MSI sensor in detecting and mapping the spatial distribution of gullies in a communal grazing landscape. *Phys. Chem. Earth* **2019**, *112*, 66–74. [CrossRef]
81. Noi, P.T.; Kappas, M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery. *Sensors* **2018**, *18*, 18. [CrossRef]
82. Jönsson, P.; Eklundh, L. TIMESAT—A program for analyzing time-series of satellite sensor data. *Comput. Geosci.* **2004**, *30*, 833–845. [CrossRef]