



Article

Evaluating the Effect of Training Data Size and Composition on the Accuracy of Smallholder Irrigated Agriculture Mapping in Mozambique Using Remote Sensing and Machine Learning Algorithms

Timon Weitkamp ^{1,2,*} and Poolad Karimi ³ ¹ Resilience BV, 6703 AA Wageningen, The Netherlands² Water Resource Management (WRM) Department, Wageningen University and Research, 6708 PB Wageningen, The Netherlands³ IHE Delft, 2611 AX Delft, The Netherlands; p.karimi@un-ihe.org

* Correspondence: timon@resiliencebv.com

Abstract: Mapping smallholder irrigated agriculture in sub-Saharan Africa using remote sensing techniques is challenging due to its small and scattered areas and heterogeneous cropping practices. A study was conducted to examine the impact of sample size and composition on the accuracy of classifying irrigated agriculture in Mozambique's Manica and Gaza provinces using three algorithms: random forest (RF), support vector machine (SVM), and artificial neural network (ANN). Four scenarios were considered, and the results showed that smaller datasets can achieve high and sufficient accuracies, regardless of their composition. However, the user and producer accuracies of irrigated agriculture do increase when the algorithms are trained with larger datasets. The study also found that the composition of the training data is important, with too few or too many samples of the "irrigated agriculture" class decreasing overall accuracy. The algorithms' robustness depends on the training data's composition, with RF and SVM showing less decrease and spread in accuracies than ANN. The study concludes that the training data size and composition are more important for classification than the algorithms used. RF and SVM are more suitable for the task as they are more robust or less sensitive to outliers than the ANN. Overall, the study provides valuable insights into mapping smallholder irrigated agriculture in sub-Saharan Africa using remote sensing techniques.

Keywords: irrigated agriculture; training data; sub-Saharan Africa; machine-learning algorithms; class imbalance



Citation: Weitkamp, T.; Karimi, P. Evaluating the Effect of Training Data Size and Composition on the Accuracy of Smallholder Irrigated Agriculture Mapping in Mozambique Using Remote Sensing and Machine Learning Algorithms. *Remote Sens.* **2023**, *15*, 3017. <https://doi.org/10.3390/rs15123017>

Academic Editor: Shawn C. Kefauver

Received: 12 May 2023

Revised: 5 June 2023

Accepted: 8 June 2023

Published: 9 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The size and composition of training samples are critical factors in remote sensing classification, as they can significantly impact classification accuracy. While sampling design is well-documented in the literature [1–5], questions remain about the optimal number of samples required, their quality, and class imbalance [6–8]. Class imbalance occurs when one or more classes is more abundant in the dataset than others, and since most machine learning classifiers try to decrease the overall error, the models are biased towards the majority class, leading to lower performances in classifying minority classes than majority classes [9]. Generally, class imbalance can be dealt with through (i) model-oriented solutions, whereby misclassifications are penalized, or where the algorithm focusses on a minority class, or (ii) data-oriented solutions, where classes are balanced by over- or undersampling [10].

Collecting a large number of quality training samples can be challenging due to limited time, access, or interpretability constraints. Practical issues and budget limitations can affect the sampling strategy, particularly in areas that are difficult to access, where rare land cover classes may be under-represented compared to more abundant classes [7,11].

Additionally, if data quality is a concern, selecting an algorithm that is less sensitive to such issues may be necessary. In the above cases it would be valuable to know how the sample size and composition affect the classification, and if additional samples are needed for increased accuracies. On the other hand, if a large sample size is already available, it may influence the choice of classifier.

These questions are even more relevant for monitoring and mapping the extent of irrigated agriculture. In particular, smallholder irrigated agriculture is often inadequately represented in datasets and policies aimed at agricultural production and irrigation development, due to informal growth and lack of government or donor involvement [12–15]. This results in an underrepresentation of smallholder irrigation in official statistics, even though smallholders provide most of the local food.

There are two general reasons for this underrepresentation. The first is the often modernistic view of what constitutes irrigation by officials and data collectors [16], in other words, large scale systems. The second reason is that African smallholder agriculture is complex, with variability in field shape, cropping systems, and timing of agronomic activities [12,17,18], often in areas that are hard to reach. Government officials and technicians that do not know about these areas will not visit them, fortifying the idea that there is no other irrigation than the large-scale systems (which are easier to reach and to recognize). Even if they do know about these systems, they might mislabel the very heterogeneous irrigated fields (i.e., many weeds) as natural vegetation.

To our knowledge, there have not been any studies yet that have investigated the effects of these biases in the training data set on classification results, and how choices made by the data collector result in changing accuracies. Choices could include oversampling irrigated agriculture because that is the class of interest, or being restricted in budget and only collecting a few samples [11]. investigated the effects of sample size on different algorithms, and we build on their ideas by including possible scenarios of how biased datasets can lead to misrepresentation.

There is ample literature on best practices regarding sampling strategies; however, these are not always followed. Although training data (TD) is often assumed to be completely accurate, it almost always contains errors [5]. These errors can come from issues with the sample design and the collection process itself and can lead to significant inaccuracies in maps created using machine learning algorithms, which can negatively impact their usefulness and interpretation [19]. It is very likely that data collection efforts in sub-Saharan Africa (SSA) are biased towards classes of interest, or heavily underestimate rare classes. That is why the main objective of this study is to investigate how different training data sizes and compositions affect the classification results of irrigated agriculture in SSA, and what the trade-offs are between cost, time, and accuracy.

This research focuses on mapping smallholder irrigation in complex landscapes in two provinces of Mozambique and explores the effects of different training data sets on the classified extent of irrigated agriculture in four scenarios: (1) Size (same ratio, smaller dataset), (2) Balance (equal numbers per class), (3) Imbalance (over- and undersampling irrigated agriculture), and (4) Mislabeling (assigning wrong class labels). To fully understand the specific effects of each type of noise source, this study uses three commonly used algorithms (RF, SVM, and ANN) in cropland mapping. This research aims to inform analysts on the effects of noise in TD on irrigated agriculture classification results.

2. Materials and Methods

Figure 1 shows the overview of the method and how the various scenarios (explained in Section 2.5) are run for the three algorithms, random forest (RF), support vector machine (SVM), and artificial neural network (ANN).

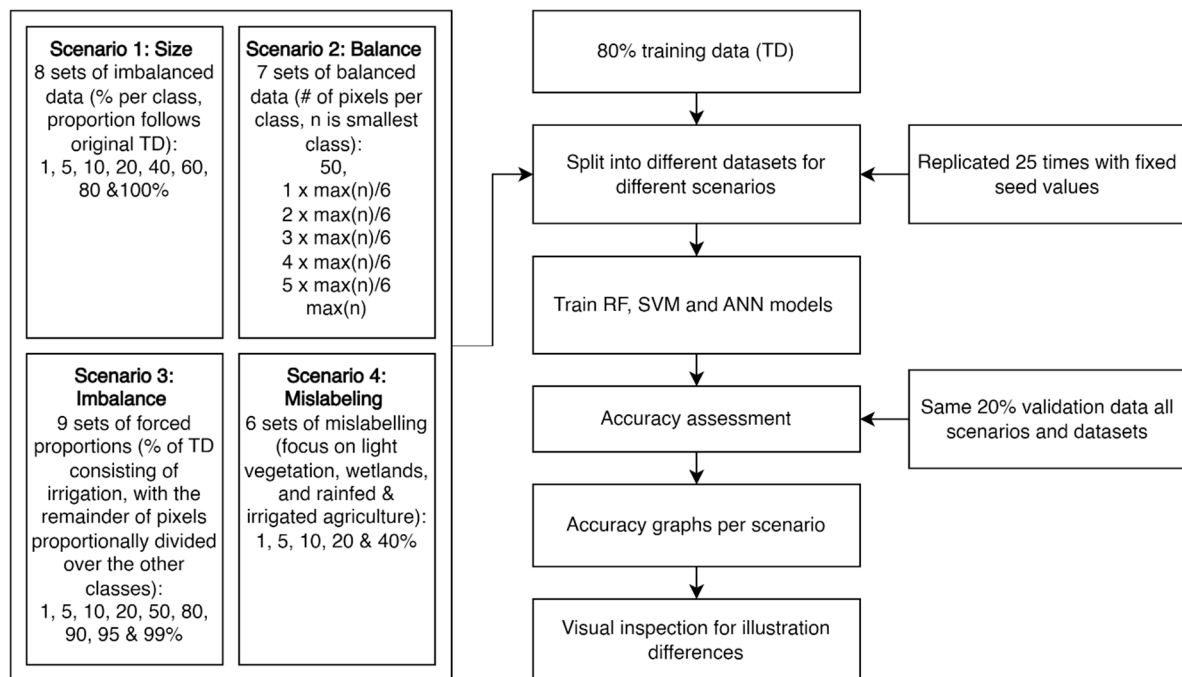


Figure 1. General overview of the methodology.

2.1. Study Area and RS Data

In this study, we compare two provinces, each with two study areas of 40×40 km (Figure 2). The two provinces are different in climate and landscape, allowing for more comparisons between models. These study areas were chosen as they contain diverse landscapes such as dense forests, wetlands, grasslands, mountains, and agriculture.

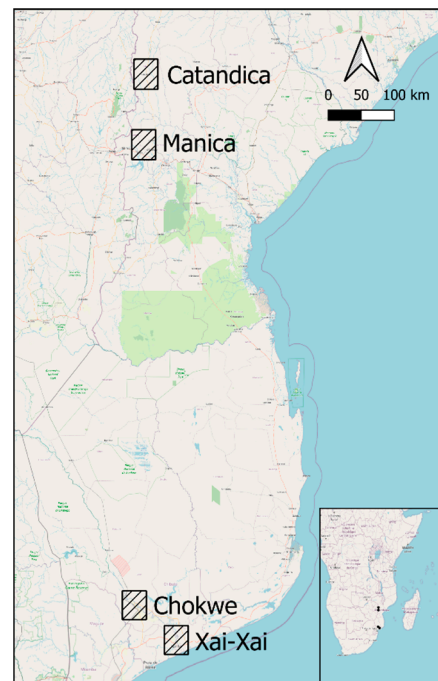


Figure 2. The four study areas, from top to bottom: Catandica and Manica in Manica province; Chokwe and Xai-Xai in Gaza province.

The following land-cover classes were mapped for this analysis (Table 1):

Table 1. Class descriptions.

Cropland Irrigated	Croplands under Management Mainly during the Dry Season
Cropland rainfed	Croplands under management mainly during the wet season
Dense vegetation	Natural vegetation comprising mainly of trees and dense undergrowth.
Light vegetation	Natural vegetation comprising of mainly low shrubs, grasses, and some trees.
Grassland	Natural vegetation of primarily grass.
Wetland	Natural vegetation that is submerged part of the year (mainly during the rainy season and first part of the dry season).
Water	Water bodies and rivers.
Built-up area	Man-made surfaces and built-up areas, including bare areas such as sand (no vegetation).

Satellite data for the four areas were collected within the Digital Earth Africa (DEA) ‘sandbox’, which provides access to Open Data Cube products in a Jupyter Notebook environment [20]. Geomedian products from Sentinel-1 and 2 were generated at a resolution of 10 m for two 6-monthly composites, representing the hydrological year from October 2019 to September 2020 [21]. The geomedian approach is a robust, high-dimensional statistic that maintains relationships between spectral bands [20,21]. Images with cloud cover exceeding 30% were filtered out in the case of Sentinel-2 data.

The normalized difference vegetation index (NDVI), bare soil index (BSI), and normalized difference water index (NDWI) were calculated using the DEA indices package for the Sentinel-2 composites [22]. In addition, the chlorophyll index red-edge (CI_{RE}) was calculated in R [23,24]. Furthermore, three second-order statistics, namely median absolute deviations (MADs), were computed using the geomedian approach: Euclidean MAD (EMAD) based on Euclidean distance, spectral MAD (SMAD) based on cosine distance, and Bray–Curtis MAD (BCMAD) based on Bray–Curtis dissimilarity, as described by [21].

Sentinel-1 data was also utilized in this study, specifically the VV and VH bands, to calculate the radar vegetation index (RVI). The use of these bands and the RVI has been documented in recent agricultural mapping studies [14,25,26]. The VV polarization data is known for its sensitivity to soil moisture, while the VH polarization data is more sensitive to volume scattering, which is influenced by vegetation characteristics and alignment. Consequently, VH data has limited potential for estimating soil moisture compared to VV data, but it exhibits higher sensitivity to vegetation [27]. The RVI has been employed in previous studies to distinguish between soil and vegetation [28,29].

To integrate all the relevant bands and indices, a comprehensive dataset was created, consisting of 18 variables (see Table 2). The specific scripts can be found on GitHub (<https://github.com/TimonWeitkamp/training-data-size-and-composition>, accessed on 11 May 2023).

Table 2. Overview of variables used.

Group	Variable	Equation
Sentinel-2	Blue	
	Green	
	Red	
	Near Infrared (NIR)	
	Red-edge 1 (RE1)	
	Red-edge 2 (RE2)	
	Shortwave Infrared 1 (SWIR1)	
	Shortwave Infrared 2 (SWIR2)	

Table 2. *Cont.*

Group	Variable	Equation
Indices S2	Normalized Difference Vegetation Index (NDVI)	$(\text{NIR} - \text{Red})/(\text{NIR} + \text{Red})$
	Normalized Difference Water Index (NDWI)	$(\text{NIR} - \text{SWIR1})/(\text{NIR} + \text{SWIR1})$
	Bare Soil Index (BSI)	$((\text{Red} + \text{SWIR1}) - (\text{NIR} + \text{Blue})) / ((\text{Red} + \text{SWIR1}) + (\text{NIR} + \text{Blue}))$
	Chlorophyll index (CI)	$(\text{NIR}/\text{Red Edge 1}) - 1$
Temporal variation	3 MADS S2	See [21,22] for more details on equations
Sentinel-1	VV	
	VH	
Indices S1	RVI	$4 \times \text{VH}/(\text{VV} + \text{VH})$

2.2. Training and Validation Samples per Scenario

Table 3 shows the number of polygons (and hectares) collected per class per study area in a clustered random strategy, supplemented with some additional irrigated pixels (purposely sampled). During the simulations, we grouped the samples based on their province to increase the total amount of training data per simulation.

Table 3. Polygon distribution and size (hectares) per area and class.

	Manica Province Catandica		Manica		Gaza Province Chokwe		Xai-Xai	
	# polygons	hectares	# polygons	hectares	# polygons	hectares	# polygons	hectares
Built-up area	10	3.4	10	5.6	10	11.5	10	18.1
Cropland irrigated	45	16.4	58	10.2	68	166	157	38.3
Cropland rainfed	34	10.9	32	7	48	40.4	19	5.8
Dense vegetation	9	148	19	104	15	12.5	9	37.2
Grassland							52	111
Light vegetation	25	89.5	20	11.3	104	187	28	26
Water			9	113	5	17.2	9	42.6
Wetland					12	144	6	27
Total	123	268.2	148	251.1	262	578.6	290	306

Of this data, the same 20% of the data per class (fixed seed number) was excluded from the training dataset intended for validation; hence, each of the results is compared with the same validation data.

This paper investigates four aspects of training data (TD) errors resulting from various sources, focusing on irrigated agriculture. The following scenarios will be explored:

Scenario 1: Size (*same ratio, smaller dataset*). In this scenario, we investigate the relationship between the amount of training data (TD) and the model's accuracy. Specifically, we want to determine whether adding more TD in the same ratio always leads to better results or if similar results can be achieved with less data.

To do this, we used eight imbalanced data sets, each with a different proportion of the original training data. The data sets ranged in size from 1% to 100% of the original dataset, with increments of 1, 5, 10, 20, 40, 60, 80, and 100%. The pixel ratio for set 8 of both provinces is shown in Table 4.

Table 4. Number of pixels in set 8 per province (size dataset).

Class	Gaza	Manica
	Set 8 (100%)	Set 8 (100%)
Built-up area	2849	1064
Irrigated agriculture	19,601	3260
Rainfed agriculture	4798	2540
Dense vegetation	6111	22,185
Grassland	10,157	-
Light vegetation	20,386	9782
Water	5504	9720
Wetland	16,582	-

Scenario 2: Balance (*equal numbers per class*). In this aspect of the study, we will examine the effect of class balance in the training data on the classification results. Simple random sampling often results in class imbalance, where rare classes are under-represented in the training set due to their smaller area. In particular, we will investigate the impact of using larger, balanced datasets on the classification performance.

We used seven sets of balanced data to achieve this, where each class has the same number of TD samples. The first set consists of 50 samples, and the remaining sets will be divided into six equal steps based on the class with the lowest abundance (i.e., the smallest class determines the step sizes). The specific sample sizes (in pixels) for each set are shown in Table 5.

Table 5. Number of pixels per set (balanced dataset).

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7
Gaza	50	508	966	1424	1882	2340	2798
Manica	50	225	400	575	750	925	1100

Scenario 3: Imbalance (*over- and undersampling irrigated agriculture*). In this scenario, we aim to investigate the effect of class imbalance caused by purposive sampling on the classification performance. Specifically, we will simulate a scenario where the proportion of samples from the class “irrigated agriculture” is increased at the cost of other classes.

To do this, we created nine sets of data, each with a different proportion of “irrigated agriculture” samples. The proportions will be 1%, 5%, 10%, 20%, 50%, 80%, 90%, 95%, and 99%. To ensure that the same total amount of training data is used in each set, the number of samples for the other classes were adjusted accordingly. The remaining training data were divided equally among the other classes, following the method described in [8] The number of samples in each class for each set is summarized in Table 6.

Table 6. Number of pixels per set (imbalanced dataset).

	Class	Set 1 (1%)	Set 2 (5%)	Set 3 (10%)	Set 4 (20%)	Set 5 (50%)	Set 6 (80%)	Set 7 (90%)	Set 8 (95%)	Set 9 (99%)
Gaza	Irrigated agriculture	202	1008	2015	4030	10,076	16,122	18,137	19,144	19,950
	Rest of the classes (7)	2850	2735	2591	2303	1439	576	288	144	29
	Total	20,152	20,153	20,152	20,151	20,149	20,154	20,153	20,152	20,153

Table 6. Cont.

	Class	Set 1 (1%)	Set 2 (5%)	Set 3 (10%)	Set 4 (20%)	Set 5 (50%)	Set 6 (80%)	Set 7 (90%)	Set 8 (95%)	Set 9 (99%)
Manica	Irrigated agriculture	54	268	535	1071	2677	4283	4819	5086	5300
	Rest of the classes (5)	1060	1017	964	857	535	214	107	54	11
	Total	5354	5353	5355	5356	5352	5353	5354	5356	5355

Scenario 4: Mislabeling (*assigning wrong class labels*). In this study, we will examine the effect of mislabeling on the classification accuracy. In smallholder agriculture SSA, class labels can be misassigned due to the heterogeneous nature of the agriculture and the potential for errors or intentional mislabeling.

To simulate this scenario, we created five sets of data, each with a different proportion of mislabeled pixels. The proportions were 1%, 5%, 10%, 20%, and 40%. The focus will be on mislabeling classes that may be considered “border cases” that are likely to be confused, rather than randomly selected classes, following [1]. These classes are irrigated agriculture, rainfed agriculture, and light vegetation. The number of misclassified pixels is shown in Table 7.

Table 7. Total number of pixels mislabeled per set for non-focus classes (irrigated and rainfed agriculture and light vegetation).

	Set 1 (1%)	Set 2 (5%)	Set 3 (10%)	Set 4 (20%)	Set 5 (40%)
Gaza	860	4299	8599	17,198	34,396
Manica	486	2428	4855	9710	19,420

2.3. Algorithm and Cross-Validation Parameter Tuning

We have used three different algorithms, namely radial support vector machines (SVM), random forests (RF), and artificial neural networks (ANN). For a description of the algorithms, we refer readers to [11,30–32]. We want to illustrate that the algorithms may interpret the data differently and lead to different classifications with different accuracies.

We used the caret package [33], which uses the free statistical software tool R (version 4.1.2) and allows for systematically comparing different algorithms and composites in a standardized method. We used rf, svmRadial, and nnet algorithms from caret for the random forest, support vector machine, and artificial neural network, respectively.

Cross-validation is a widely used method for evaluating the performance of machine learning algorithms and models. In cross-validation, the data is divided into multiple folds or subsets, typically of equal size. The algorithm is trained on one subset and tested on the other subsets, so each subset is used for testing exactly once. The algorithm’s performance is then evaluated based on the average performance across all the folds.

Spatial K-fold cross-validation is a variation of the traditional cross-validation approach that considers the spatial relationships between the samples in the dataset [34]. The spatial k-folds method divides the data into k subsets, with each subset consisting of samples that are spatially close to each other. This is particularly useful in remote sensing, where the spatial relationships between the samples are important in understanding the underlying patterns in the data. In this study, we used spatial k-fold cross-validation.

2.4. Classifications and Replications

To ensure the accuracy and reliability of our models, we conducted 25 iterations of all steps for each of the three algorithms using the same seed numbers. By replicating the process, we could account for the variability in accuracies that may depend on the specific training data sets used in each run. This allowed us to evaluate the robustness

and generalizability of the models and determine whether they were sensitive to specific training data points and seed numbers or whether they were more robust and generalizable to the study area.

We created various sample sizes and compositions by using random subsampling from the complete sample set, with different seed values. To decrease computation time, we used the `caret::train()` function and included all variables in the model rather than using forward feature selection of the variables.

Figure 3 displays the range of model parameter values per scenario, training data set, and province based on the overall accuracy. The range of values used by the same algorithms across different seed values and scenarios demonstrates the inherent randomness in the model results, even with the same training data. Some parameter values, such as the `mtry` value of 2 for RF and the decay and size values for ANN, consistently show higher preference across all datasets. However, sigma from SVM exhibits little overlap between the provinces and scenarios. These findings suggest that parameter tuning is highly recommended for SVM and ANN while less necessary for rf, as evident from the lack of clear patterns in the results—similar to what [35] also found.

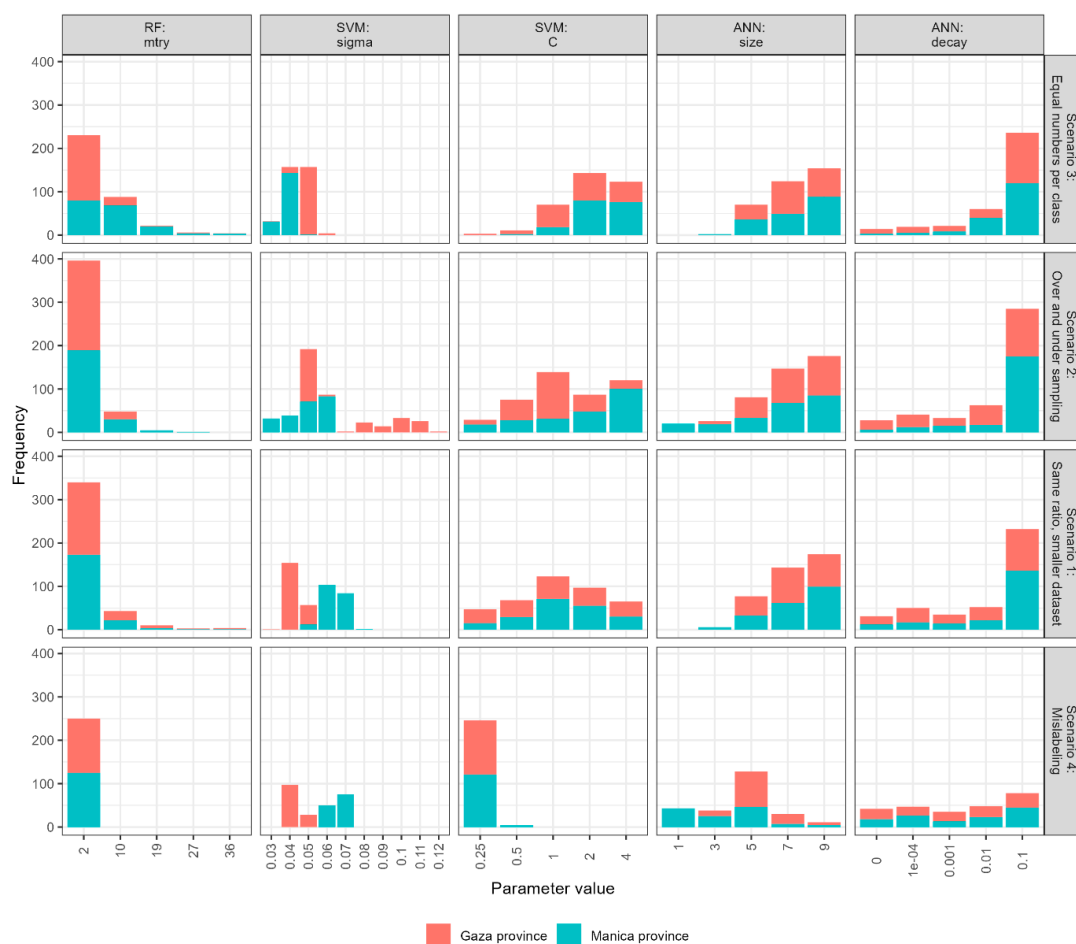


Figure 3. Parameter values and how often a model uses that value per algorithm, per scenario (dataset).

2.5. Accuracy Assessment

We calculated the overall accuracy and the user's and producer's accuracies using the same validation dataset for each iteration (Table 8).

Table 8. Sample sizes per class used for accuracy assessment.

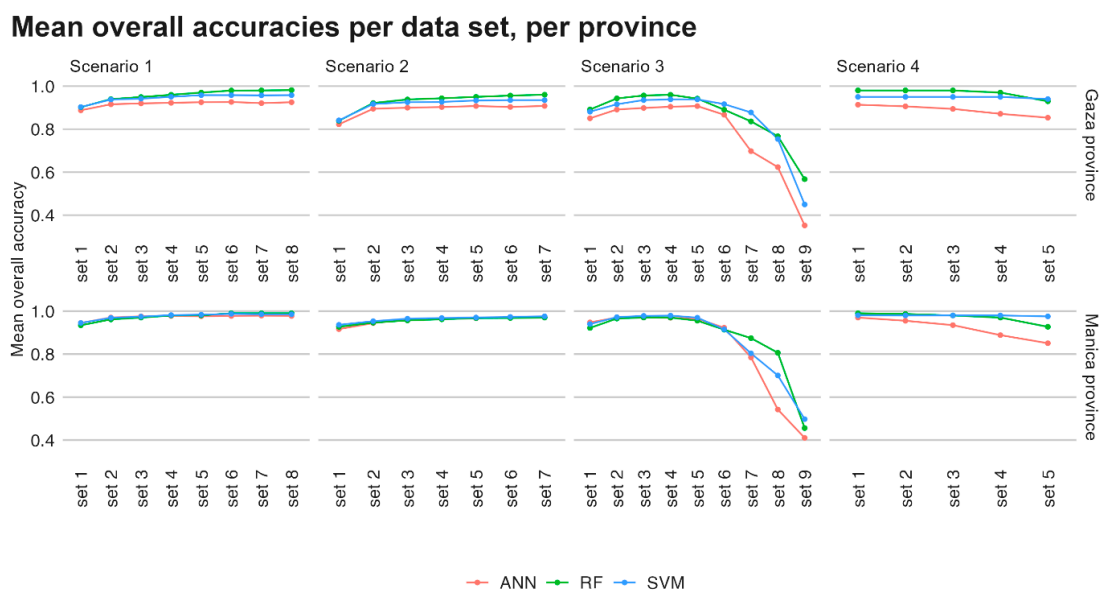
	Gaza	Manica
Built-up area	668	252
Irrigated agriculture	4936	823
Rainfed agriculture	1227	607
Dense vegetation	1496	5577
Grassland	2536	-
Light vegetation	5132	2428
Water	1339	2452
Wetland	4165	-

3. Results

The four scenarios (Tables 3–6) were designed to demonstrate the impact of training data composition on accuracy, based on possible design and collection errors. Firstly, each scenario’s mean overall accuracy per dataset is presented, separated by the province to account for varying climates and agricultural regions. Then, a closer examination of the classification of irrigated agriculture within each scenario is conducted, using the user and producer accuracies.

3.1. The Overall Accuracy of All Scenarios

Figure 4 summarizes the mean overall accuracy of the three classification methods, per scenario and study area. In scenarios 1 (same class ratio, but smaller) and 2 (equal number of pixels per class), high accuracy plateaus of greater than 90% are achieved within the first two sets (5% of total and 508/225 pixels per class, respectively), with similar results across all algorithms. In scenario 3, which involves over- and undersampling of the “irrigated agriculture” class, the accuracy starts high and peaks at sets 3 and 4. However, depending on the algorithm used, it decreases to less than 30–60% in Gaza and 40–50% in Manica when more than three quarters of the dataset contains a single class. Scenario 4, which involves mislabeling, shows high accuracy with the first sets (1–5% mislabeling), with the SVM algorithm remaining particularly stable, while the other two algorithms drop by only five percentage points.

**Figure 4.** Mean overall accuracies per algorithm, dataset, province, for each scenario.

The overall accuracy is mainly affected by the majority classes and hides considerable variation of individual runs. Thus, we will also investigate the classification results of the irrigated agriculture class by using user and producer accuracies.

3.2. Class Specific Accuracies per Scenario

3.2.1. Scenario 1: Same Ratio, Smaller Dataset

Figure 5 compares the accuracies of irrigated agriculture between Gaza and Manica using different algorithms, for scenario 1. Generally, larger datasets (set 8) show higher accuracies and less variation in values per dataset than smaller datasets, although there are still differences between the algorithms and study areas.

1. Size - Same ratio but smaller dataset

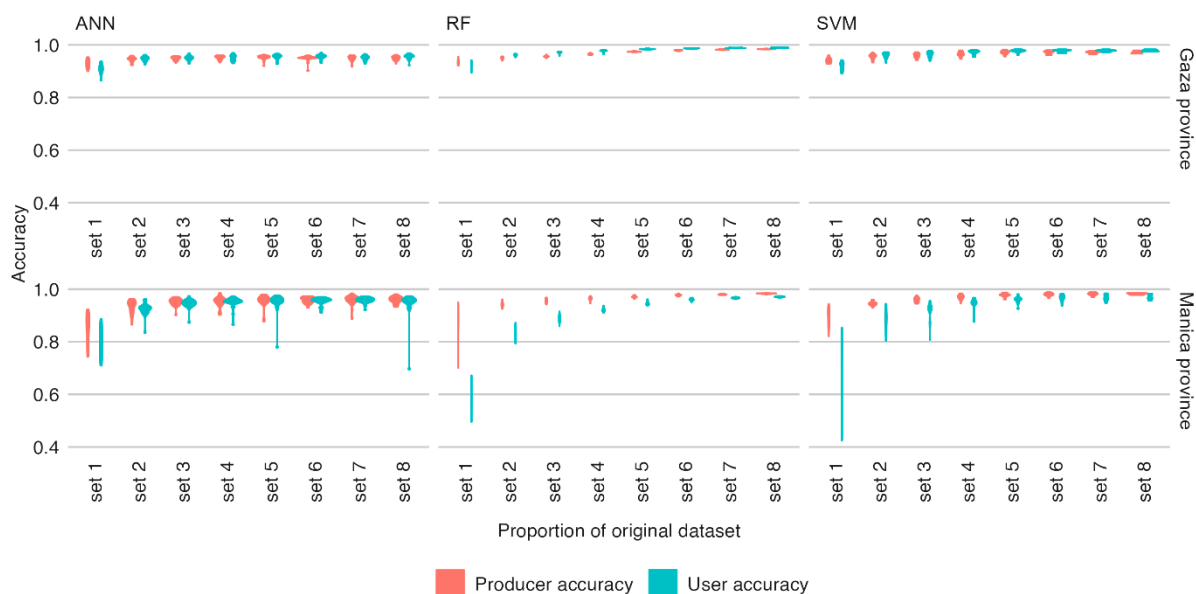


Figure 5. Distribution of user and producer accuracy irrigated agriculture for each algorithm and dataset, per province, for scenario 1: size.

In Gaza, the more homogeneous study area, the RF algorithm has the lowest accuracy spread and the highest accuracy values, whereas the SVM and ANN have more spread and slightly lower accuracies. The three algorithms are quite stable, with set 2 already leading to comparable results as set 8, which is 10–20 times larger. For each algorithm, the user and producer accuracies are in the same range, indicating that “irrigated agriculture” (user), as well as other classes (producer), are accurately classified. The accuracies are also similar to the mean overall accuracies.

In Manica, which is more heterogeneous, the user and producer accuracies start low and increase until a plateau of ~95% is reached after the fifth set with all algorithms. The most extensive spread in values can be found with ANN in all sets and both accuracies, followed by SVM in the user accuracy, whereas RF shows the least spread in values. Set 1 (the smallest dataset) has the lowest accuracies with the largest spread with all algorithms. However, ANN still has high accuracy (around 80%). It also reaches the plateau the fastest, suggesting that ANN performs well on smaller datasets, albeit with a larger spread, indicating sensitivity to the specific dataset used. The user accuracy is generally lower than the producer accuracy for RF and SVM, at least in the first few sets, indicating that these models were less able to identify “irrigated agriculture” (user), but better at identifying other classes (producer). This could be due to the models not being exposed to enough “irrigated agriculture” samples in the training phase or the models overfitting other classes, meaning they can classify those classes well but not the “irrigated agriculture” class. The producer’s accuracy is in line with the mean overall accuracy, whereas the user’s is less so.

3.2.2. Scenario 2: Equal Numbers per Class

While the producer accuracy is higher than the user accuracy in Gaza, it is the other way around in Manica (Figure 6). In Gaza, this indicates that the models are not very good at identifying the class of interest (irrigated agriculture) to the user, but they are very good at identifying other classes. In Manica, the models are very good at identifying the class of interest (irrigated agriculture) to the user but not as good at identifying other classes.

2. Balance - Equal numbers per class

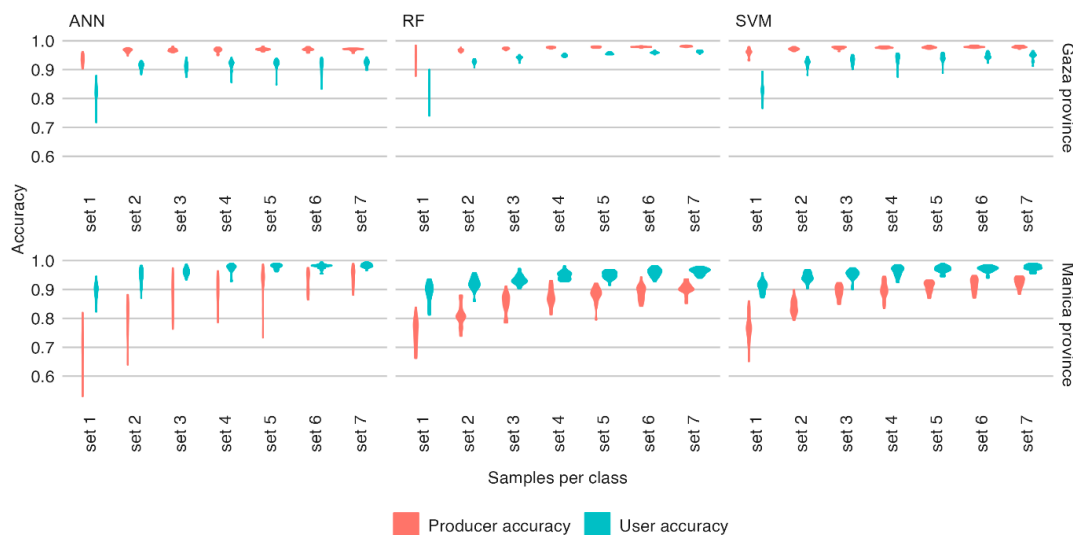


Figure 6. Distribution of user and producer accuracy irrigated agriculture for each algorithm and dataset, per province, for scenario 2: equal numbers per class.

In Gaza, most of the producer accuracy values are well above 95%, indicating that almost all the training data samples have been correctly classified. The user accuracies, although high, show more spread in values and remain lower (only the last sets reach 95%), indicating that there is a slight overestimation of irrigated agriculture, especially when the training data contains fewer irrigated agriculture pixels (first few sets). Excluding set 1, RF has the least spread in values, followed by SVM. ANN seems to have the most difficulty in consistent classifications, even as the total number of pixels increases.

In Manica, there is an overall increase in class-specific accuracies with an increasing sample size of irrigated agriculture across all three algorithms (Figure 6). The spread in accuracies in the models with the most irrigated agriculture pixels (set 7) is less than those with fewer samples (set 1), suggesting more robust classifications. However, there is not much difference between the last four sets. ANN shows the largest spread in producer accuracies between the algorithms and starts with the lowest accuracies, while RF and SVM show less spread. Although ANN showed the largest spread, it also achieved the highest accuracies (between 90–95%), followed by RF and SVM with slightly lower accuracies (85–95%). The user accuracies of the three algorithms are more similar and mostly above 90% accuracy, with ANN having the smallest (set 7) and largest (set 1) spread and the highest accuracies, followed by RF and SVM with slightly lower accuracies and larger spreads.

3.2.3. Scenario 3: Over- and Undersampling

Scenario 3, as shown in Figure 7, reveals that the user and producer accuracies are similar around sets 3 and 4, which contain between 10–20% of the “irrigated agriculture” class. This composition is similar to those of the training dataset in Gaza and Manica, which are 22% and 6%, respectively. The producer accuracy remains high until set 4, after which it drops rapidly as the proportion of “irrigated agriculture” increases. The user accuracy is the opposite and increases until set 4, after which it reaches 100% accuracy. This is not

surprising, as most of the map will be classified as “irrigated agriculture,” meaning the validation data will be correct for that class. The other classes will be less present in the later sets, resulting in a low producer accuracy.

3. Imbalance - Over and under sampling

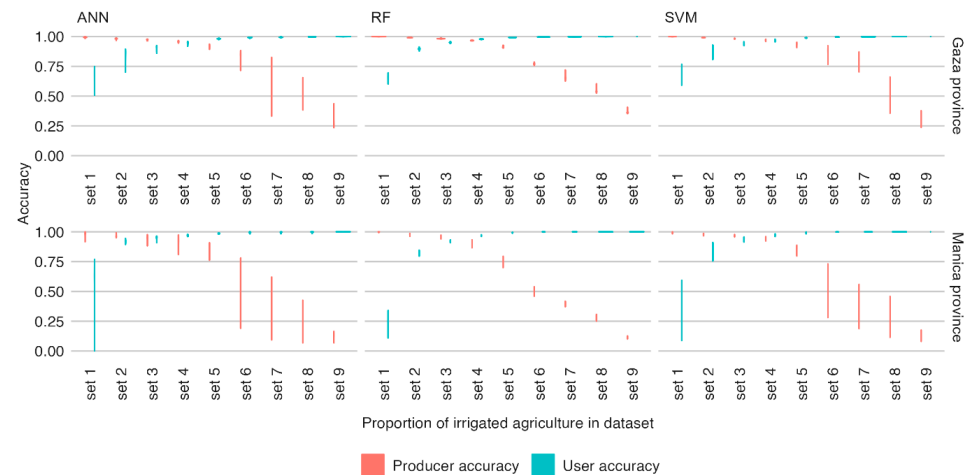


Figure 7. Distribution of user and producer accuracy irrigated agriculture for each algorithm and dataset, per province, for scenario 3: over- and undersampling.

The RF algorithm shows the least spread in both user and producer accuracy. ANN and SVM have larger spreads in producer than user accuracy, and user accuracy spread is small after sets 2/3. Producer accuracy spread starts small but increases with each set for these two algorithms.

3.2.4. Scenario 4: Mislabeling Irrigated, Rainfed, and Light Vegetation

Scenario 4 (Figure 8) reveals that in Gaza, the SVM algorithm’s accuracies remain high in all five sets (over 95%), with only a slight decrease in accuracy and minimal spread in values. The RF algorithm follows this trend but dips slightly lower in set 5. ANN has the largest downward trend and the most spread in accuracy values.

4. Mislabeling - assigning wrong class labels

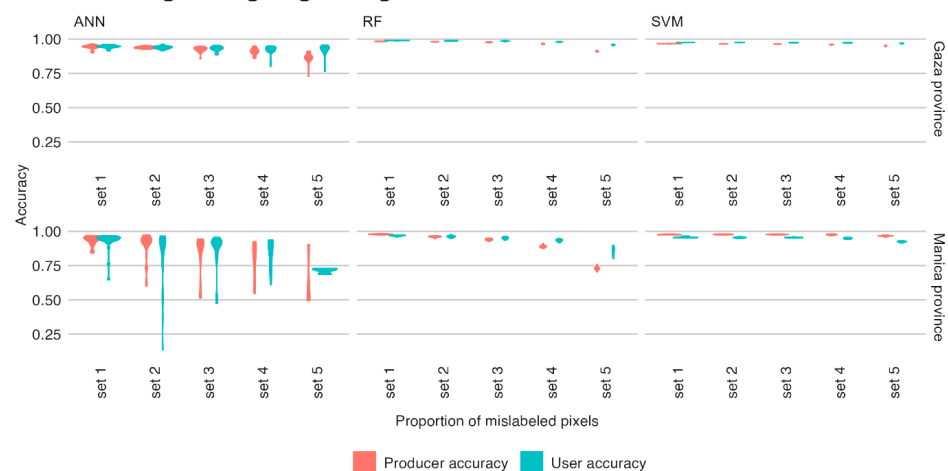


Figure 8. Distribution of user and producer accuracy of irrigated agriculture for each algorithm and dataset, per province, for scenario 4: mislabeling.

In Manica, as seen in Gaza, the SVM algorithm performs best with stable and high (over 95%) accuracies. The RF algorithm starts high but drops to 75–85% accuracy in the

last set, with slightly more spread in values. The ANN algorithm has the largest spread and a larger downward trend.

3.3. Visual Inspection

In this section, we present a visualization of the level of agreement among models for classifying irrigated agriculture in the Chokwe area. The images depict areas with varying degrees of green and red, with darker shades indicating higher agreement or disagreement among models (referred to as agreement maps), respectively. Specifically, the darkest green shade corresponds to areas where 25 models agreed on the classification of the pixel as irrigated agriculture, while the dark red shade indicates a classification by only one model. In cases where no red or green shades are present, it means that the pixel was classified as a different class other than irrigated agriculture. We have chosen to display only the first and last sets per scenario to illustrate the extremes.

3.3.1. Scenario 1: Same Ratio, Smaller Dataset

Figure 9 presents a comparison between the results of set 1 (1% of the data) and set 8 (100% of the data) for scenario 1, and Figure 10 shows the area in hectares for each agreement value, split over the northern and southern region of the Limpopo River. Our analysis reveals that set 8 identifies a substantially higher amount of irrigated agriculture compared to set 1, particularly in the southern region of the Limpopo River, which encompasses the Chokwe Irrigation Scheme (CIS). In contrast, the northern bank consists of rain-fed agriculture and farmer-led irrigation. Set 1 performs poorly in identifying irrigated agriculture in this region with the ANN and SVM algorithms; the RF algorithm shows more irrigated agriculture. We also see that set 8 has more hectares in the upper regions of agreement than set 1, which has more hectares in the lower agreement values. In other words, more area is more confidently classified in set 8 than in set 1. The figure also shows that there is a large increase in hectares with the 25-agreement value (all models agree), especially the northern region, which almost doubles in size with all three algorithms.

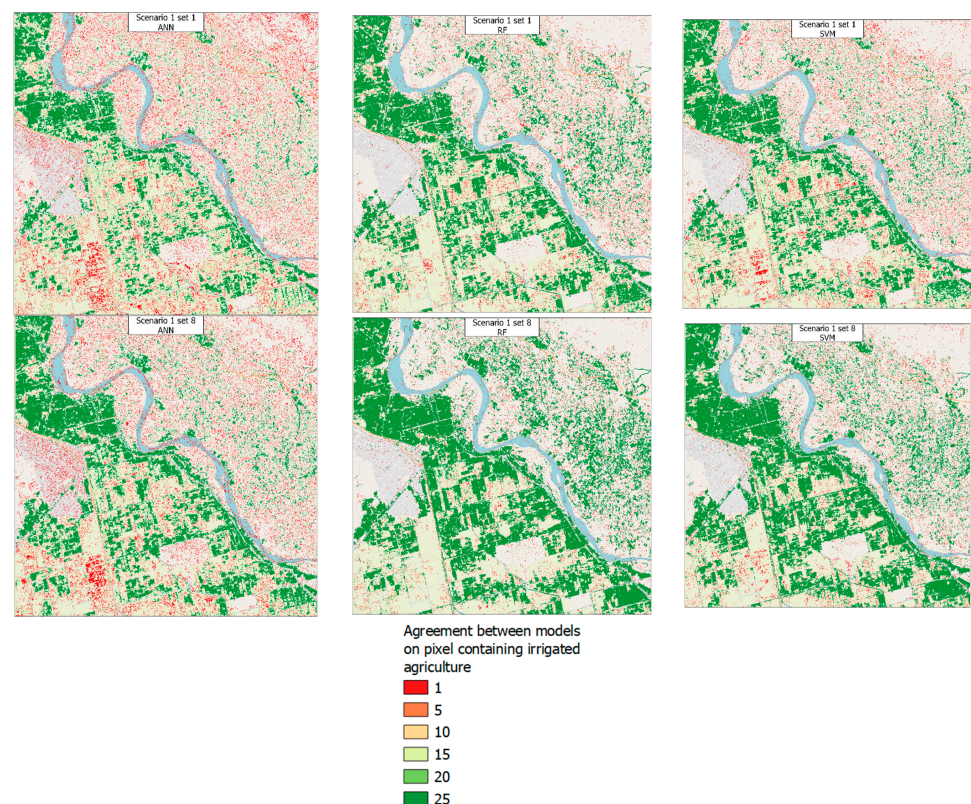


Figure 9. Scenario 1 agreement maps.

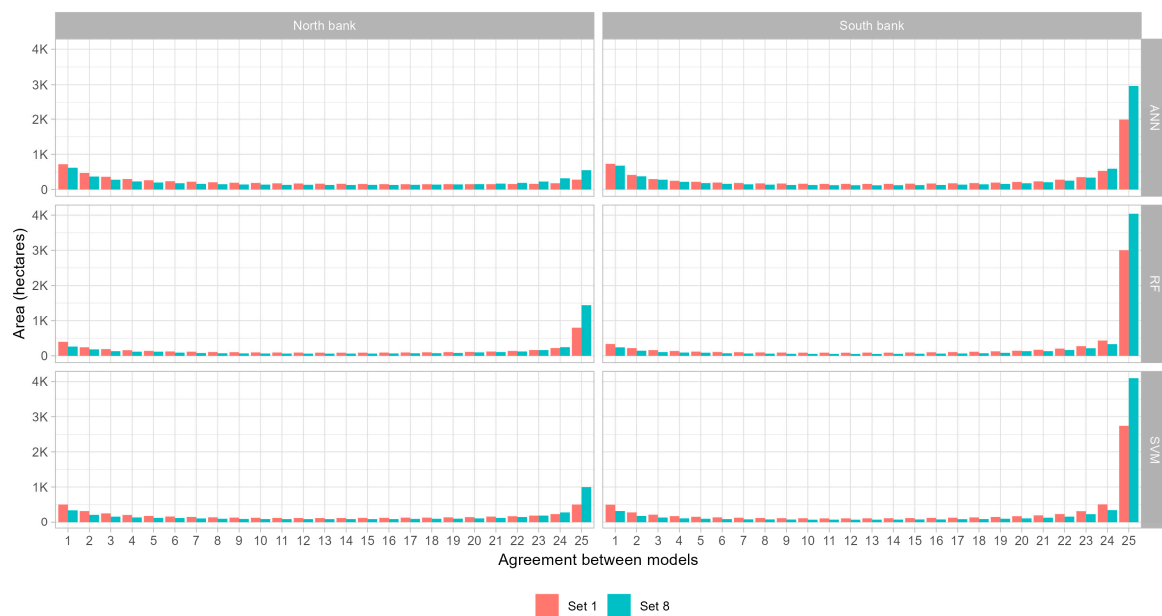


Figure 10. Scenario 1: number of hectares per *agreement* from the different maps of Chokwe, split over north and south of the Limpopo River to highlight smallholder and conventional irrigation systems.

Furthermore, we observed differences in the performance of the algorithms. The ANN algorithm identified considerably less irrigated agriculture than the RF and SVM algorithms, which demonstrated similar performances (Figure 10).

3.3.2. Scenario 2: Equal Numbers per Class

Scenario 2 (Figure 11), where each class has the same number of pixels, shows more significant differences between the smallest and largest datasets than scenario 1 (Figure 9). Set 1 underclassifies the CIS and shows limited irrigation agriculture on the northern bank. The red pixels, where only a few models classify irrigated agriculture, mostly correspond to individual trees or small groups of trees. In contrast, set 7 presents a more balanced map with fewer red areas and larger clusters of irrigated agriculture.

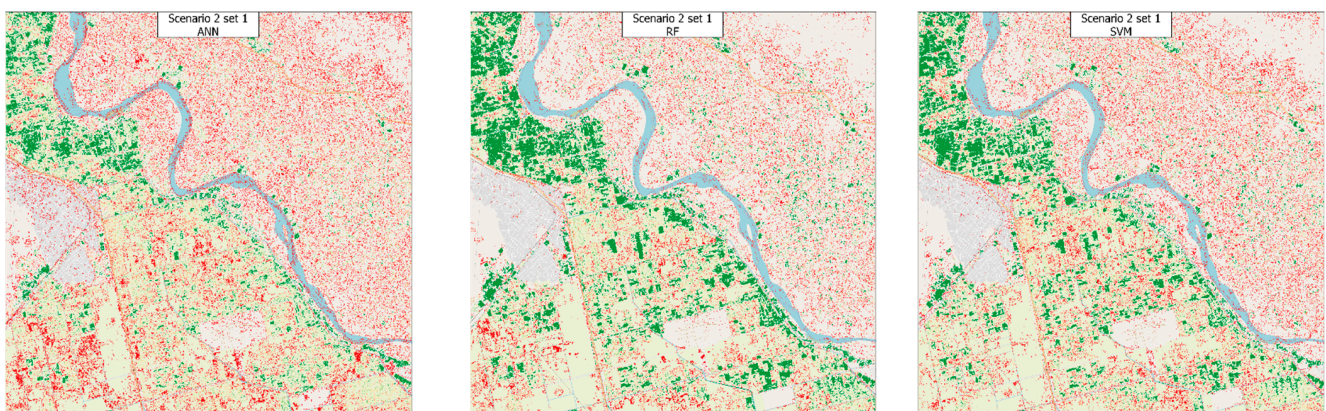


Figure 11. Cont.

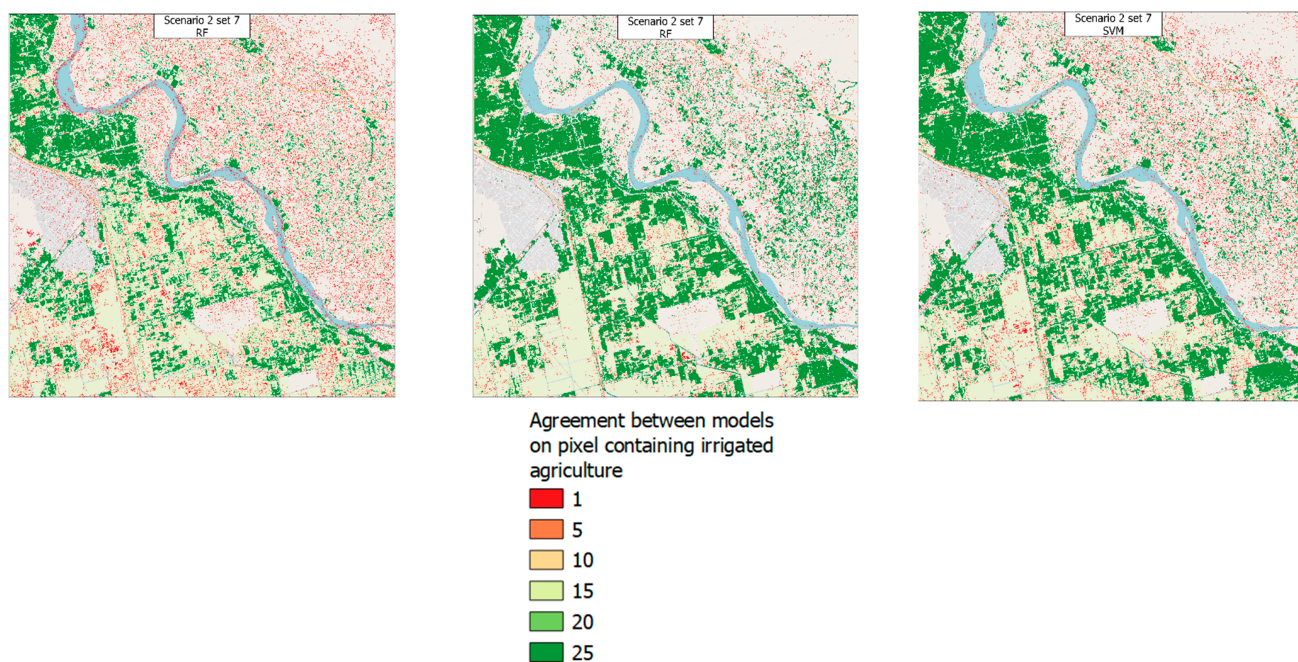


Figure 11. Scenario 2 agreement maps.

The RF and SVM maps are similar in both sets, while ANN shows fewer areas classified as irrigated agriculture, similar to scenario 1. Additionally, ANN misclassifies the natural vegetation on the Limpopo banks as irrigated agriculture in both sets.

Figure 12 shows that the smaller datasets lead to less agreement between the models; there are more hectares of irrigated agriculture only identified by one or a couple of the models. Compared to Figure 10, set 7 shows fewer hectares of irrigated agriculture when using a balanced dataset near the 25-agreement range, whereas the lower agreement range is similar.

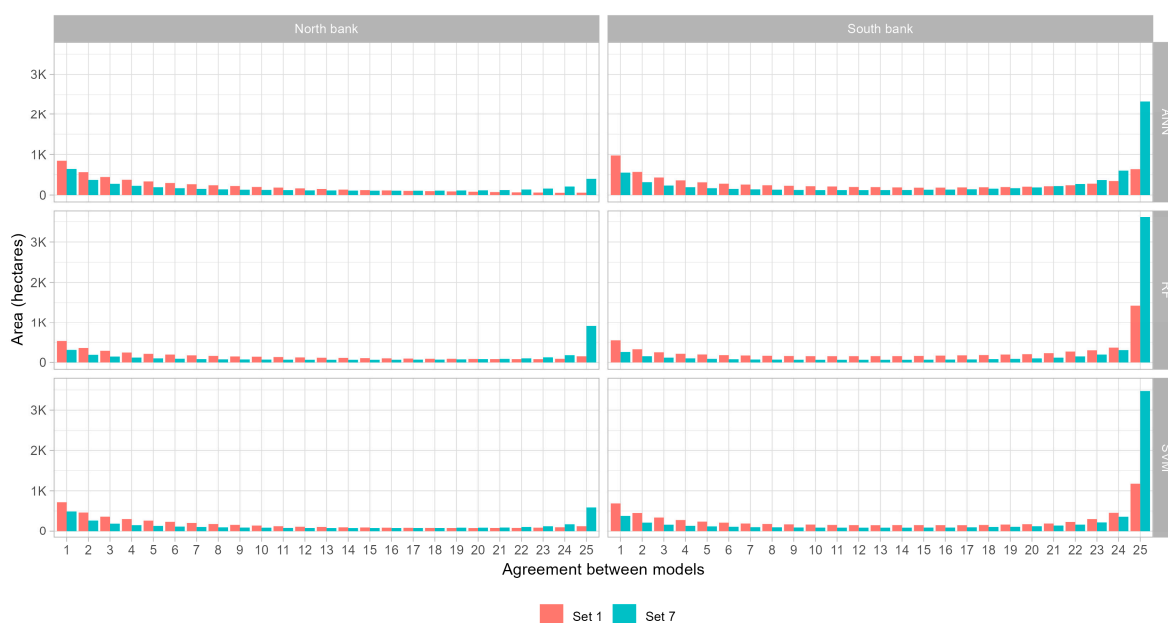


Figure 12. Scenario 2: number of hectares per agreement from the different maps of Chokwe, split over north and south of the Limpopo River to highlight smallholder and conventional irrigation systems.

3.3.3. Scenario 3: Over- and Undersampling

Scenario 3 highlights the impact of over- and undersampling of irrigated agriculture, where set 1 has only 1% of the pixels classified as irrigated agriculture, while set 9 has 99% (Figure 13). As expected, having very little training data for irrigated agriculture results in limited classification of that class, while having almost only class-specific training data leads to cleaner maps with fewer red areas on the north bank (at least for RF and SVM).

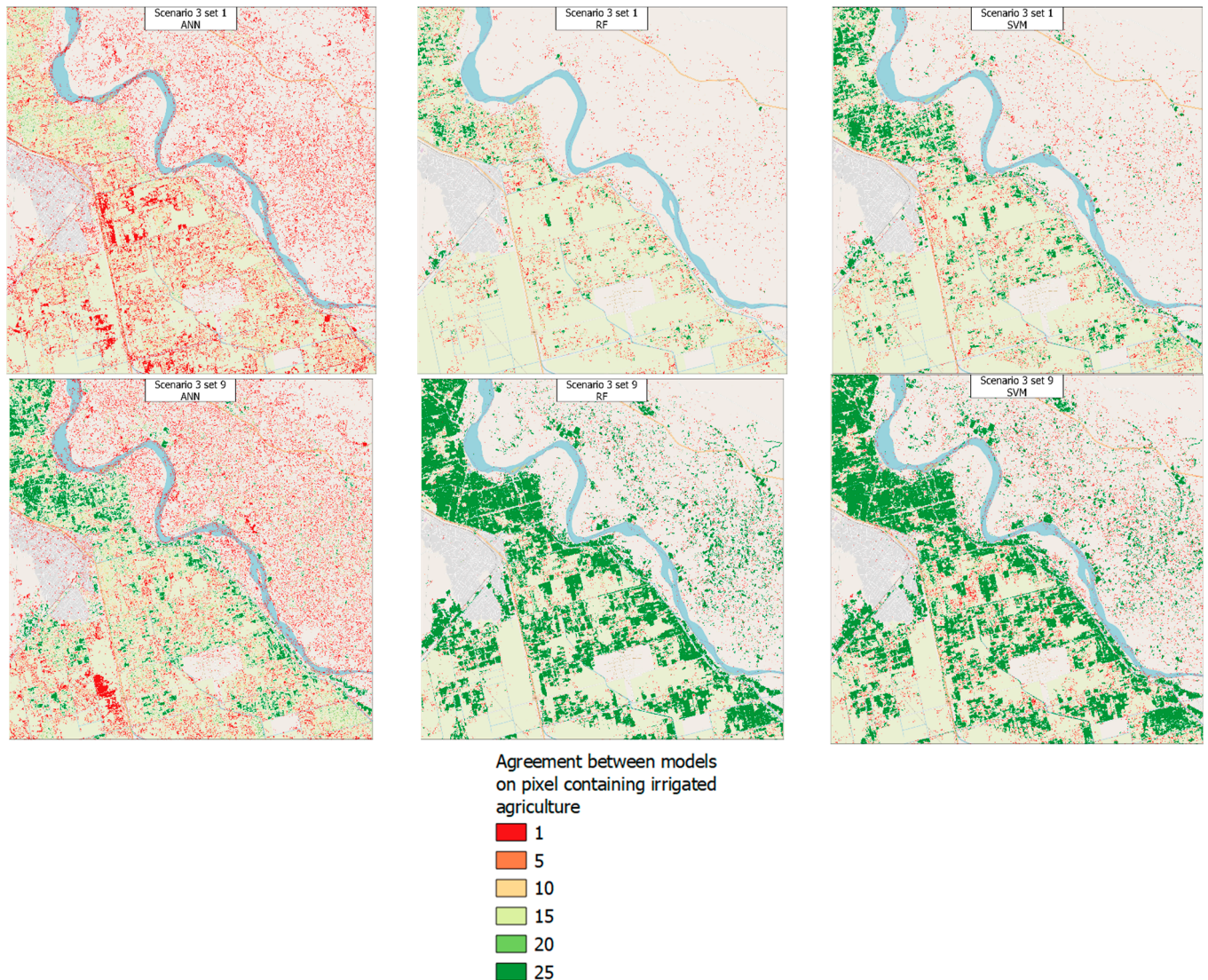


Figure 13. Scenario 3 agreement maps.

Comparing the algorithms, we observe that ANN classifies more irrigated agriculture in set 1 than the other two algorithms, but there is minimal agreement among the 25 models (no green areas present in set 1). Set 3 using ANN shows more irrigated agriculture, but still less than the other two algorithms. With less data (set 1), RF and SVM are less similar, but in set 9, they become more similar again.

Figure 14 reiterates these findings; additionally, the total area of 25-agreement of set 5 is still lower than in the previous two scenarios, showing that oversampling the class of interest is not beneficial.

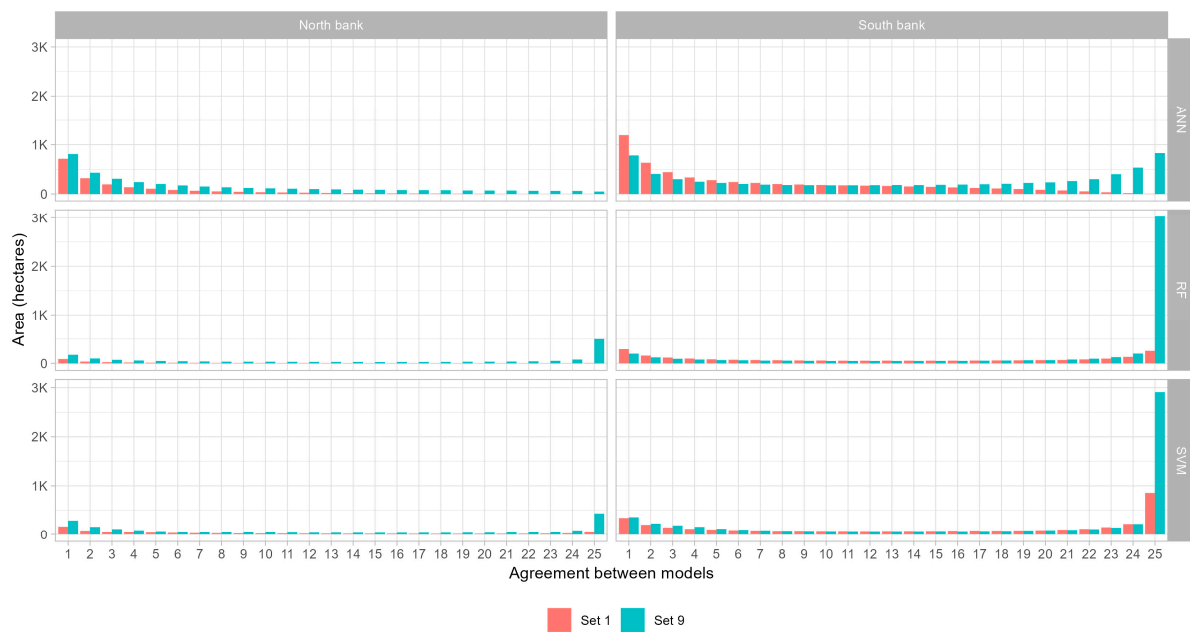


Figure 14. Scenario 3: number of hectares per agreement from the different maps of Chokwe, split over north and south of the Limpopo River to highlight smallholder and conventional irrigation systems.

3.3.4. Scenario 4: Mislabeling Irrigated, Rainfed, and Light Vegetation

In Figure 15, we compare Scenario 4 set 4 (with 40% misclassification) with scenario 1 set 8 (with 0% misclassification) for reference. Scenario 4 set 4 (Figures 15 and 16) shows that almost as much irrigated agriculture is classified on the north bank as the south bank, with all three algorithms, compared to the other scenarios. At the same time, there is less irrigated agriculture in the CIS, with more emphasis on heterogeneous areas for classifying irrigated agriculture.

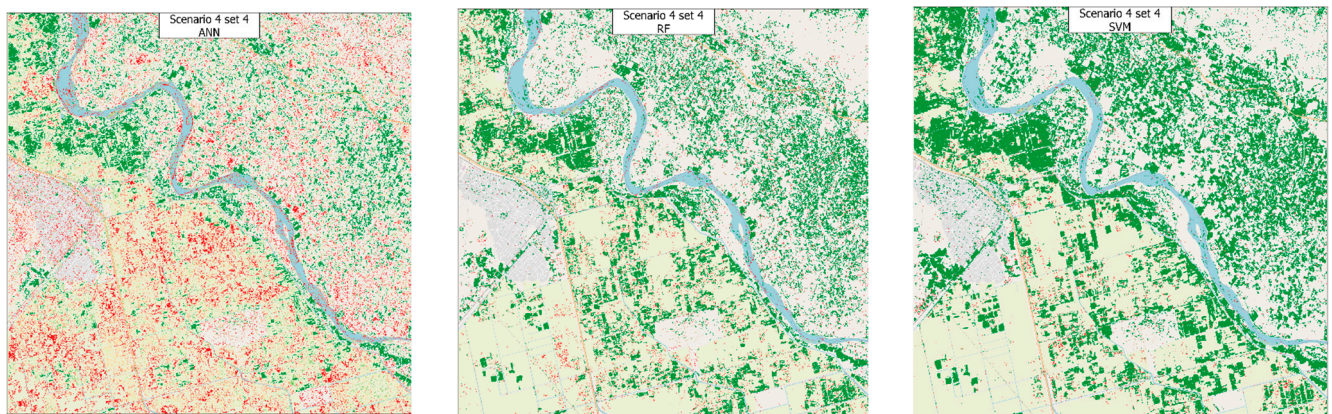


Figure 15. Cont.

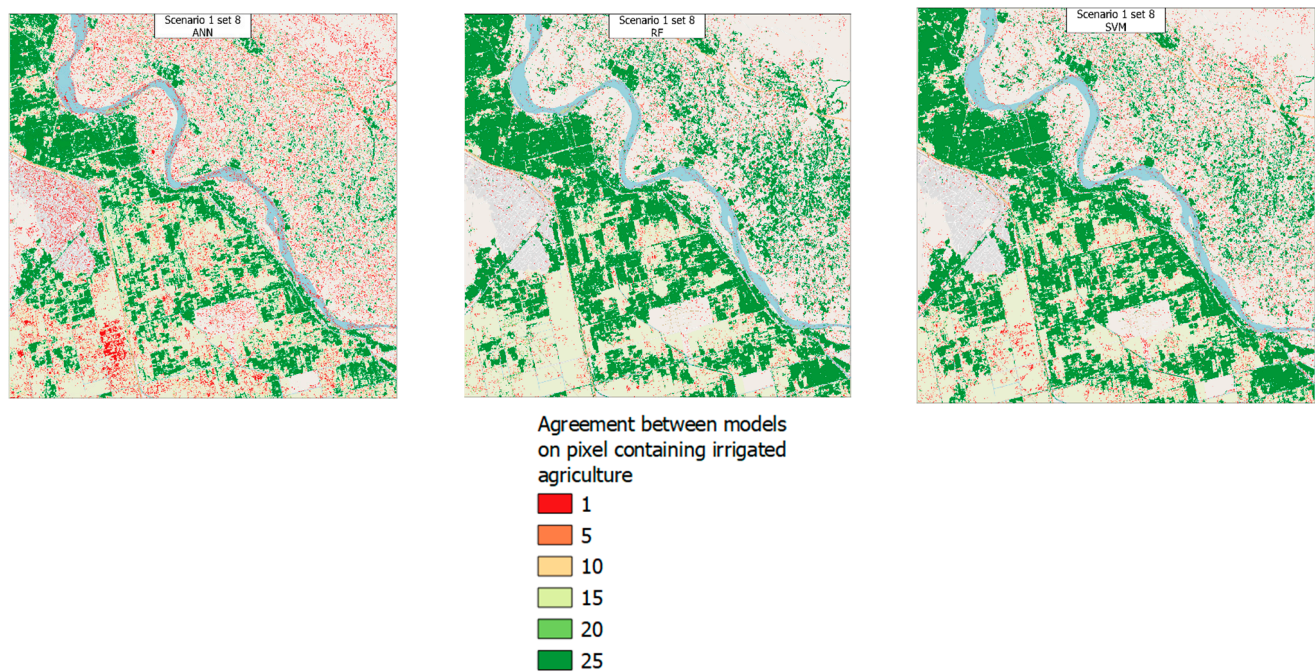


Figure 15. Scenario 4 agreement maps.

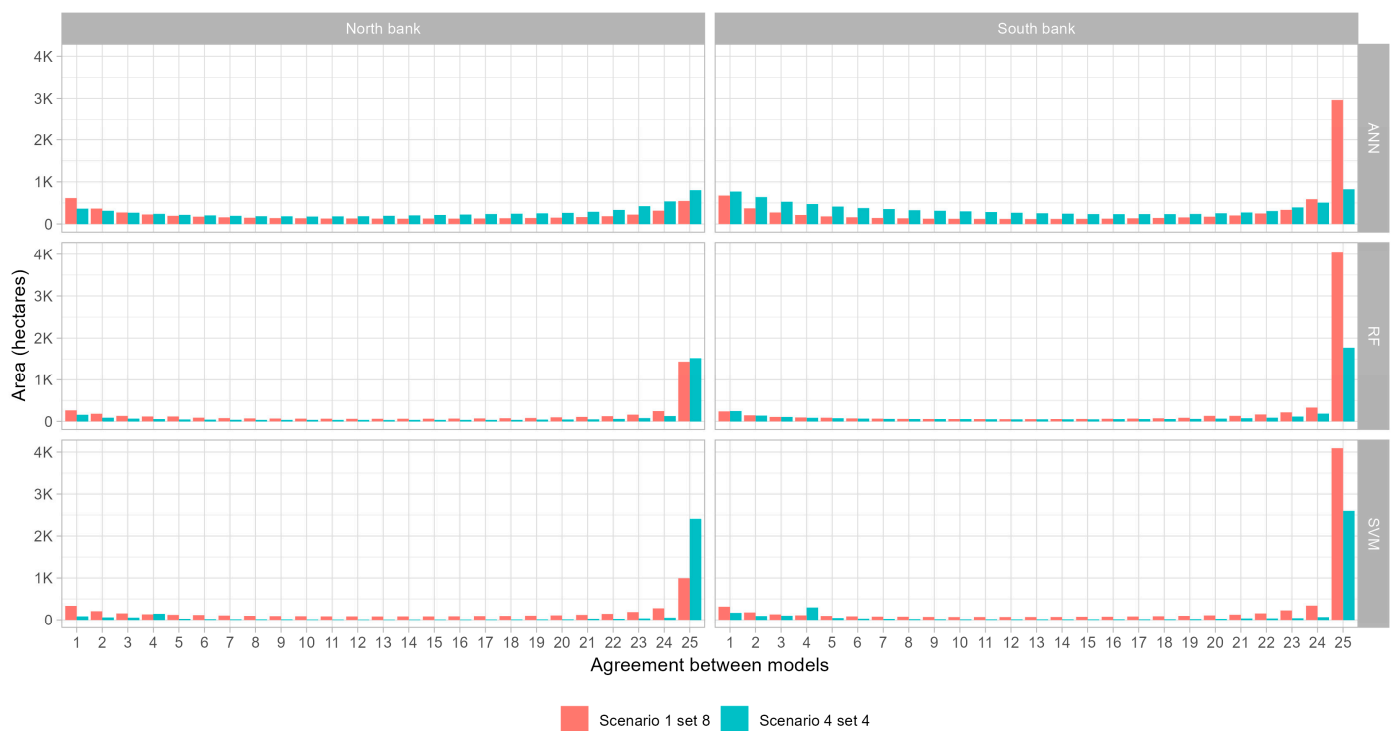


Figure 16. Scenario 4: number of hectares per agreement from the different maps of Chokwe, split over north and south of the Limpopo River to highlight smallholder and conventional irrigation systems.

As in all previous scenarios, the ANN algorithm classifies the least area as irrigated agriculture (Figure 16), followed by RF. The SVM algorithm classifies the most irrigated agriculture.

4. Discussion

The results of this study align with previous research by [11], which found that larger sample sizes lead to improved classifier performance and that increasing the sample set size after a certain point did not substantially improve the classification accuracy. Scenarios 1 and 2 in this research show that larger datasets improve overall classification results, but not by much. This plateauing of overall accuracy is not unexpected, as when classifications reach very high overall accuracy, there is little potential for further increases. Our study is also in line with the results of [11], in that user and producer accuracies continued to increase with larger sample sizes, indicating that larger sample sizes are still preferable to smaller sizes, even with similar overall accuracy results.

A large spread in accuracy means that the specific results depend more on the dataset that is used for that classification than other factors. For example, the SVM algorithm in Manica in Scenario 1 resulted in a user accuracy of just above 40%, but also 85%. By chance, any of the two could have become the final classification; if it was the 85% classification, one would think enough data was collected for the study, whereas the other sets show that higher accuracies are possible, with less spread in values. The lower spread in values also indicates a more stable model which can generalize more. It also means that the specific dataset used for the classification is less important, as similar results can be expected from any random subset, also seen in Section 3.3.

Scenario 1, where eight datasets ranging in size from 1% to 100% of the original dataset were used, shows that larger training datasets lead to higher user and producer accuracy with less spread in values (Figure 5). The size of set 5 in Manica falls between sets 3 and 4 of Gaza (40% vs. 10–20%, respectively), which are also the sets after which the accuracies plateau in Gaza. This corresponds to ~1300 pixels of irrigated agriculture for Manica and ~1900–3900 for Gaza. This reinforces the statement that larger training data sets are preferable over smaller sets but that there is an optimum after which accuracies only marginally increase at the cost of more computing time and, effectively, more resources are ‘lost’ collecting that data in the first place. To find out if enough data is collected for a classification of irrigated area, researchers and practitioners can use this subsetting method to evaluate if different iterations yield the same, stable results, or if additional resources should be put towards more field data collection.

Scenario 2 also examines the impact of data size on classification performance, but with equal numbers of samples per class, spread over seven sets. Similar to scenario 1, larger datasets generally result in higher user and producer accuracies (Figure 6). However, this scenario highlights differences in the performance of the classifiers in the two study areas. In scenario 1, the results of both study areas followed similar patterns but with different accuracy values. In this scenario, however, the user and producer accuracy trends are reversed, depending on the study area. In Gaza, the user’s accuracy is consistently lower than the producer’s, whereas in Manica, the user’s accuracy is consistently higher than the producer’s. Manica also shows a larger spread in values for both user and producer accuracy.

This trend reversal suggests that the models in Gaza are better able to classify the non-irrigated agriculture classes than the irrigated agriculture class, indicating a more generalized model. Conversely, the Manica models can better classify the irrigated agriculture class than the non-irrigated agriculture classes, indicating a less generalized model. As all classes have the same number of pixels per dataset within the same study area, the complexity of the landscape likely plays a role in this difference. The two provinces generally have different landscapes (flat vs. mountain), climate (little vs. much rainfall) and consequently, different agricultural practices, with different field sizes (larger vs. small) and shapes (regular vs. irregular). It is worth noting that, even though Gaza has twice the number of pixels as Manica, sets 1 are the same size in both cases, and 3 of Gaza and 7 of Manica are similar in size. However, even for these sets with similar sizes, Gaza has higher producer accuracies, and Manica has higher user accuracies.

Scenario 3, where irrigated agriculture is vastly over- and undersampled in nine sets ranging from 1% to 99%, shows a peak in overall accuracy around sets 3 and 4 (Figure 4, 10% and 20% irrigated agriculture in the dataset). These two sets reflect the ‘true’ composition of the dataset, which was found in the field. When irrigated agriculture is underrepresented (sets 1 and 2, 1% and 5%), the overall accuracy is not much lower. This is because the other majority classes have a greater impact on the overall accuracy. As more irrigated agriculture is present in the training datasets (sets 5 to 9, 50–99%), the other classes decrease in size, and irrigated agriculture becomes the majority class. The high user accuracy (Figure 7) indicates that any irrigated agriculture in the validation set is correctly classified (not surprising as all pixels are classified as such). However, the reverse is that the producer accuracy is extremely low (many of the pixels are wrongly classified as irrigated agriculture instead of a different class).

Scenario 4, where similar classes are mislabeled on purpose in five sets from 1% to 40% mislabeling, shows a decrease in overall accuracy (Figure 4) for ANN and only a minor decrease in the last set for RF. SVM does not seem to be affected, possibly because the support vectors used for distinguishing the different classes do not change much between the sets, indicating that SVM is less sensitive to data set compositions.

The user and producer accuracies (Figure 8) also show that SVM can handle this imbalance, perhaps because it uses the same support vectors to distinguish the different classes in all the sets. Adding more data will not help the algorithm, as that data is not near the separation planes between classes. RF is similarly stable, except for the last set, which also shows a larger spread in accuracy values. The user accuracy is also higher than the producer’s, which comes from slowly oversampling irrigated agriculture (among other classes). The ANN has many difficulties with changing compositions, as seen from the large spread in values and decreased accuracies. Overall, RF and SVM seem to handle this mislabeled data well.

The results of the study demonstrate the importance of the dataset and algorithm selection in accurately classifying irrigated agriculture in remote sensing data. Visual inspection reveals that different areas are classified as irrigated agriculture depending on the dataset and algorithm used. In some cases, the models prioritize farmer-led irrigated areas over more conventional large-scale irrigated areas, but the latter is generally classified more accurately. The amount of data used and the balance between classes also have a significant impact on the accuracy of classification, with too few data or imbalanced data resulting in underestimation of the extent of farmer-led irrigation, and too much noise resulting in overestimation. The RF and SVM algorithms are found to be more robust with noisy data than the ANN algorithm. Although the maps do not distinguish between farmer-led irrigation and large-scale irrigation, our knowledge of the area enables us to interpret the maps in terms of these different types of irrigation.

Generally, there are many oversampling and undersampling strategies which have not been tested. The focus of this study was not to find the best method to deal with imbalanced data, but to illustrate what imbalanced data does with the final results.

Overall, ANN showed high results but with a large spread in all scenarios and study areas. The RF and SVM showed results similar to each other, depending on the scenario’s dataset and study area, resulting in higher accuracies with lower spreads. Both are recommended for mapping irrigated agriculture. The large spread in ANN shows that it may be suitable for detecting irrigated agriculture, but only in certain circumstances—when there is much data (scenario 1 final sets), and the landscape is more homogeneous (Gaza, all scenarios). Nevertheless, the random chance of high or low accuracies is higher with ANN than with RF and SVM (i.e., larger spread), indicating that the specific dataset used in modelling is more important for ANN than the other two algorithms.

According to [31], the training sample size and quality can have a greater impact on classification accuracy than the choice of algorithm. As a result, differences in accuracy between datasets within the same algorithm should be more pronounced than those between different algorithms. This is supported by scenarios 1, 2, and 3, where the algorithms

show similar trends and values but exhibit greater variability within datasets. Scenario 3 demonstrates that user and producer accuracies may cross over, but the differences between datasets are still more significant than those between algorithms. However, scenario 4 is less conclusive, since there is little variation in the high accuracies of the RF and SVM algorithms across all sets, with some variation in Manica. At the same time, ANN shows dissimilar trends and greater differences between sets compared to the other two algorithms.

5. Conclusions

The results of this study indicate that larger sample sizes generally lead to higher user and producer accuracies. However, there is an optimum after which accuracies only marginally increase at the cost of more computing time and collection effort (Scenario 1). We also show that the models trained on Gaza were better at the classification of all classes (i.e., a more generalized model) than those trained on Manica (Scenario 2). In other words, the more homogeneous landscape of Gaza lead to models that could generally classify all classes, whereas models of the more heterogeneous Manica were overfitting towards irrigated agriculture, even though all classes had the same number of pixels in the training data sets. Scenarios 3 and 4 show that the field data collected should reflect the actual landscape composition and that class labels can bias towards heterogeneous areas (i.e., no oversampling of irrigated agriculture or mislabeling), and that random forest and support vector machine are more suitable for classifying irrigated agriculture than the artificial neural network, as they are less sensitive to the specific dataset.

This study provides valuable insights for practitioners and researchers mapping irrigated agriculture in sub-Saharan Africa by means of remote sensing techniques. It highlights the importance of carefully considering sample size and composition when collecting and using data. African smallholder agriculture is complex, with variability in field shape, cropping systems, and timing of agronomic activities. Based on this study, to accurately predict such smallholder irrigated agriculture, we recommend to:

- Ensure that training data represents the area being classified and includes sufficient samples to achieve high accuracy. This can be done best using a random sampling design. Although perfect data is desirable, models (RF and SVM) can tolerate some noise.
- Evaluate multiple algorithms when classifying data, as different algorithms may perform better or worse depending on the specific characteristics of the data being classified.
- Interpret classification results carefully, as accuracies alone may not correctly represent the classification performance. Visual inspection and further interpretation are needed to understand the results and potential limitations of the classification fully.
- Perform multiple simulations with different subsets of the data to estimate if the training data yields robust results (i.e., minimal variation in accuracies between sets), which can indicate that sufficient data has been collected.

Author Contributions: Conceptualization, T.W. and P.K.; methodology, T.W. and P.K.; formal analysis, T.W.; writing—original draft preparation, T.W.; writing—review and editing, P.K.; visualization, T.W.; supervision, P.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by International Development Research Centre (IDRS), grant number (PROJECT ID) 109039 and The APC was funded by Resilience BV.

Data Availability Statement: The irrigation maps at different spatial scales produced in this study and scripts used are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict 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.

References

1. Foody, G.; Pal, M.; Rocchini, D.; Garzon-Lopez, C.; Bastin, L. The Sensitivity of Mapping Methods to Reference Data Quality: Training Supervised Image Classifications with Imperfect Reference Data. *Int. J. Geo-Inf.* **2016**, *5*, 199. [\[CrossRef\]](#)
2. Foody, G.M. Sample Size Determination for Image Classification Accuracy Assessment and Comparison. *Int. J. Remote Sens.* **2009**, *30*, 5273–5291. [\[CrossRef\]](#)
3. Foody, G.M.; Mathur, A.; Sanchez-Hernandez, C.; Boyd, D.S. Training Set Size Requirements for the Classification of a Specific Class. *Remote Sens. Environ.* **2006**, *104*, 1–14. [\[CrossRef\]](#)
4. 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. [\[CrossRef\]](#)
5. Stehman, S.V.; Foody, G.M. Key Issues in Rigorous Accuracy Assessment of Land Cover Products. *Remote Sens. Environ.* **2019**, *231*, 111199. [\[CrossRef\]](#)
6. Collins, L.; McCarthy, G.; Mellor, A.; Newell, G.; Smith, L. Training Data Requirements for Fire Severity Mapping Using Landsat Imagery and Random Forest. *Remote Sens. Environ.* **2020**, *245*, 111839. [\[CrossRef\]](#)
7. 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. [\[CrossRef\]](#)
8. 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. [\[CrossRef\]](#)
9. Ebrahimi, H.; Mirbagheri, B.; Matkan, A.A.; Azadbakht, M. Effectiveness of the Integration of Data Balancing Techniques and Tree-Based Ensemble Machine Learning Algorithms for Spatially-Explicit Land Cover Accuracy Prediction. *Remote Sens. Appl. Soc. Environ.* **2022**, *27*, 100785. [\[CrossRef\]](#)
10. Douzas, G.; Bacao, F.; Fonseca, J.; Khudinyan, M. Imbalanced Learning in Land Cover Classification: Improving Minority Classes' Prediction Accuracy Using the Geometric SMOTE Algorithm. *Remote Sens.* **2019**, *11*, 3040. [\[CrossRef\]](#)
11. 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. [\[CrossRef\]](#)
12. Beekman, W.; Veldwisch, G.J.; Bolding, A. Identifying the Potential for Irrigation Development in Mozambique: Capitalizing on the Drivers behind Farmer-Led Irrigation Expansion. *Phys. Chem. Earth Parts A/B/C* **2014**, *76–78*, 54–63. [\[CrossRef\]](#)
13. Veldwisch, G.J.; Venot, J.-P.; Woodhouse, P.; Komakech, H.C.; Brockington, D. Re-Introducing Politics in African Farmer-Led Irrigation Development: Introduction to a Special Issue. *Water Altern.* **2019**, *12*, 12.
14. Venot, J.-P.; Bowers, S.; Brockington, D.; Komakech, H.; Ryan, C.; Veldwisch, G.J.; Woodhouse, P. Below the Radar: Data, Narratives and the Politics of Irrigation in Sub-Saharan Africa. *Water Altern.* **2021**, *14*, 27.
15. Woodhouse, P.; Veldwisch, G.J.; Venot, J.-P.; Brockington, D.; Komakech, H.; Manjichi, A. African Farmer-Led Irrigation Development: Re-Framing Agricultural Policy and Investment? *J. Peasant Stud.* **2017**, *44*, 213–233. [\[CrossRef\]](#)
16. de Bont, C. Modernisation and African Farmer-Led Irrigation Development: Ideology, Policies and Practices. *Water Altern.* **2019**, *12*, 23.
17. Bégué, A.; Arvor, D.; Bellon, B.; Betbeder, J.; de Abelleira, D.; PD Ferraz, R.; Lebourgeois, V.; Lelong, C.; Simões, M.; Verón, S.R. Remote Sensing and Cropping Practices: A Review. *Remote Sens.* **2018**, *10*, 99. [\[CrossRef\]](#)
18. Izzi, G.; Denison, J.; Veldwisch, G.J. *The Farmer-Led Irrigation Development Guide: A What, Why and How-to for Intervention Design*; World Bank: Washington, DC, USA, 2021.
19. Elmes, A.; Alemohammad, H.; Avery, R.; Caylor, K.; Eastman, J.; Fishgold, L.; Friedl, M.; Jain, M.; Kohli, D.; Laso Bayas, J.; et al. Accounting for Training Data Error in Machine Learning Applied to Earth Observations. *Remote Sens.* **2020**, *12*, 1034. [\[CrossRef\]](#)
20. DEA. DEA GeoMAD. Available online: [https://docs.digitalearthafrika.org/en/latest/data_specs/GeoMAD_specs.html#Triple-Median-Absolute-Deviations-\(MADs\)](https://docs.digitalearthafrika.org/en/latest/data_specs/GeoMAD_specs.html#Triple-Median-Absolute-Deviations-(MADs)) (accessed on 6 September 2022).
21. Roberts, D.; Dunn, B.; Mueller, N. Open Data Cube Products Using High-Dimensional Statistics of Time Series. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; IEEE: Valencia, Spain, 2018; pp. 8647–8650.
22. Wellington, M.J.; Renzullo, L.J. High-Dimensional Satellite Image Compositing and Statistics for Enhanced Irrigated Crop Mapping. *Remote Sens.* **2021**, *13*, 1300. [\[CrossRef\]](#)
23. Gitelson, A.A.; Viña, A.; Ciganda, V.; Rundquist, D.C.; Arkebauer, T.J. Remote Estimation of Canopy Chlorophyll Content in Crops. *Geophys. Res. Lett.* **2005**, *32*, L08403. [\[CrossRef\]](#)
24. Segarra, J.; Buchailot, M.L.; Araus, J.L.; Kefauver, S.C. Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications. *Agronomy* **2020**, *10*, 641. [\[CrossRef\]](#)
25. Abubakar, G.A.; Wang, K.; Shahtahmasebi, A.; Xue, X.; Belete, M.; Gudo, A.J.A.; Mohamed Shuka, K.A.; Gan, M. Mapping Maize Fields by Using Multi-Temporal Sentinel-1A and Sentinel-2A Images in Makarfi, Northern Nigeria, Africa. *Sustainability* **2020**, *12*, 2539. [\[CrossRef\]](#)
26. Gella, G.W.; Bijker, W.; Belgiu, M. Mapping Crop Types in Complex Farming Areas Using SAR Imagery with Dynamic Time Warping. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 171–183. [\[CrossRef\]](#)
27. Gao, Q.; Zribi, M.; Escorihuela, M.; Baghdadi, N.; Segui, P. Irrigation Mapping Using Sentinel-1 Time Series at Field Scale. *Remote Sens.* **2018**, *10*, 1495. [\[CrossRef\]](#)

28. Jennewein, J.S.; Lamb, B.T.; Hively, W.D.; Thieme, A.; Thapa, R.; Goldsmith, A.; Mirsky, S.B. Integration of Satellite-Based Optical and Synthetic Aperture Radar Imagery to Estimate Winter Cover Crop Performance in Cereal Grasses. *Remote Sens.* **2022**, *14*, 2077. [[CrossRef](#)]
29. Mandal, D.; Kumar, V.; Ratha, D.; Dey, S.; Bhattacharya, A.; Lopez-Sanchez, J.M.; McNairn, H.; Rao, Y.S. Dual Polarimetric Radar Vegetation Index for Crop Growth Monitoring Using Sentinel-1 SAR Data. *Remote Sens. Environ.* **2020**, *247*, 111954. [[CrossRef](#)]
30. Abdolrasol, M.G.M.; Hussain, S.M.S.; Ustun, T.S.; Sarker, M.R.; Hannan, M.A.; Mohamed, R.; Ali, J.A.; Mekhilef, S.; Milad, A. Artificial Neural Networks Based Optimization Techniques: A Review. *Electronics* **2021**, *10*, 2689. [[CrossRef](#)]
31. 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](#)]
32. Thanh Noi, P.; Kappas, M. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors* **2017**, *18*, 18. [[CrossRef](#)]
33. Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [[CrossRef](#)]
34. Meyer, H.; Reudenbach, C.; Hengl, T.; Katurji, M.; Nauss, T. Improving Performance of Spatio-Temporal Machine Learning Models Using Forward Feature Selection and Target-Oriented Validation. *Environ. Model. Softw.* **2018**, *101*, 1–9. [[CrossRef](#)]
35. Phalke, A.R.; Özdoğan, M.; Thenkabail, P.S.; Erickson, T.; Gorelick, N.; Yadav, K.; Congalton, R.G. Mapping Croplands of Europe, Middle East, Russia, and Central Asia Using Landsat, Random Forest, and Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 104–122. [[CrossRef](#)]

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