

## Article

# Bayesian Aggregation Improves Traditional Single-Image Crop Classification Approaches

Ivan Matvienko <sup>1</sup>, Mikhail Gasanov <sup>2,\*</sup>, Anna Petrovskaia <sup>2</sup>, Maxim Kuznetsov <sup>2</sup>, Raghavendra Jana <sup>3</sup>, Maria Pukalchik <sup>4</sup> and Ivan Oseledets <sup>2,5</sup>

<sup>1</sup> Yandex LLC, 119021 Moscow, Russia

<sup>2</sup> Center for Artificial Intelligence Technology, Skolkovo Institute of Science and Technology, 121205 Moscow, Russia

<sup>3</sup> Center for Agro Technologies, Skolkovo Institute of Science and Technology, 121205 Moscow, Russia

<sup>4</sup> Public Joint-Stock Company (PJSC) Sberbank of Russia, 117997 Moscow, Russia

<sup>5</sup> Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, 119991 Moscow, Russia

\* Correspondence: mikhail.gasanov@skoltech.ru

**Abstract:** Accurate information about growing crops allows for regulating the internal stocks of agricultural products and drawing strategies for negotiating agricultural commodities on financial markets. Machine learning methods are widely implemented for crop type recognition and classification based on satellite images. However, field classification is complicated by class imbalance and aggregation of pixel-wise into field-wise forecasting. We propose here a Bayesian methodology for the aggregation of classification results. We report the comparison of class balancing techniques. We also report the comparison of classical machine learning methods and the U-Net convolutional neural network for classifying crops using a single satellite image. The best result for single-satellite-image crop classification was achieved with an overall accuracy of 77.4% and a Macro F1-score of 0.66. Bayesian aggregation for field-wise classification improved the result obtained using majority voting aggregation by 1.5%. We demonstrate here that the Bayesian aggregation approach outperforms the majority voting and averaging strategy in overall accuracy for the single-image crop classification task.

**Keywords:** crop classification; unbalanced classes problem; pixel-wise aggregation



**Citation:** Matvienko, I.; Gasanov, M.; Petrovskaia, A.; Kuznetsov, M.; Jana, R.; Pukalchik, M.; Oseledets, I. Bayesian Aggregation Improves Traditional Single-Image Crop Classification Approaches. *Sensors* **2022**, *22*, 8600. <https://doi.org/10.3390/s22228600>

Academic Editor: Lorena Parra

Received: 29 September 2022

Accepted: 7 November 2022

Published: 8 November 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

Crop identification from satellite images is an important problem in precision agriculture. Possessing accurate information about growing crops makes it possible to regulate agricultural products' internal stocks and to draw strategies for negotiating agricultural commodities on financial markets. Recent developments in remote sensing and data processing techniques have enabled both scientists and practical workers to simplify the process of crop identification.

The growing use of satellite imagery at very high spatial and temporal resolution allows land managers to obtain extensive data on how the land is used. Different methods for mapping crop patterns by classifying multi-source and multi-temporal data have been proposed and tested by using time series from different satellites in recent years [1]. However, most of these studies demonstrated these methods having certain drawbacks, which were mainly related to the high cost of data collection for time-series and the cloud cover problem, which is commonly faced in different regions. For example, snow cover may stay for more than four months in northern countries, such as Canada or Russia. For some countries, such as India, monsoon season with its dense cloud cover could be the reason for the absence of satellite data. This essentially means that sometimes only one or two cloud-free images can be obtained from satellites and analyzed during the vegetation period. Therefore, an adequate single-image crop classification approach should be used in this case.

Although ML and DL classification algorithms are used to classify crops, the processed classification results' aggregation toward the field level is not considered in the previous articles. It is also worth noting that most works use multi-temporal data for training. However, how to conduct one-show multi-spectral image learning for crop classification needs to be clarified. The objective of this study was to examine resampling and aggregation techniques for ML and DL models. In this work, we compared ML and DL models, such as convolutions neural networks (CNNs), for crop type recognition. The Sentinel-2 satellite image was used as input data for classification models. We compared classical aggregation methods of pixel-wise classification and also suggested a new approach to aggregate results towards field-wise ones using a Bayesian strategy.

This paper is organized as follows. Section 2 describes the related ML and DL approaches for crop classification, current methods of resampling to tackle class imbalance problems, and recent Bayesian methods applications. Section 3 describes the datasets, satellite image preprocessing and spectral indices generation, ML and DL models used for classification, classical resampling methods, and Bayesian aggregation proposed in this paper. Section 4 presents the results of crop classification, such as overall accuracy and macro-F1 score, and comparison of aggregation methods. Section 5 presents the conclusions.

The main contributions of this paper are as follows:

- We demonstrated the performance of ML models and the U-Net neural network to address crop classification problems;
- We compared resampling techniques to tackle the problem of high imbalance of classes;
- We proposed a new Bayesian aggregation strategy of pixel-wise classification.

## 2. Related Works

Several major approaches are used to tackle the problem of crop classification based on satellite images. For the last decade, classical machine learning (ML) was the most popular and powerful instrument used for these purposes. Many authors have already studied crop type classification using traditional statistical or machine learning methods including Random Forest [2–4], Discriminant analysis [5], k-Nearest Neighbors, Extreme Learning Machine [6], Maximum Likelihood Classification [7,8], CART Decision Trees [4,9] and Support Vector Machine (SVM) [7,10–14]. Recently, multi-spectral Sentinel-2 images and SAR Sentinel-1 data were combined for the crop mapping problem supported with the Random Forest model in Germany; for major crops, the F1-score was equal to 0.8 [15].

Class imbalance has a serious effect on crop type recognition. For example, class imbalance explains 40% of the accuracy variability [16]. Recently, class balancing techniques have been compared, showing improved results using SMOTE-based methods [17]. Moreover, researchers have studied the impact of phenology-based methods on the problem of class imbalance, while absence samples for classification were automatically generated using dynamic time warping [18].

A number of researchers reported on the use of neural networks to tackle the problem of crop classification. For example, deep convolutional neural network [19–21] and LSTM recurrent neural networks [21] were successfully used for crop recognition. High-resolution Planet satellite tiles and time series of Sentinel-2 tiles were used for the segmentation of crop types in Africa and Germany using 2D U-Net + CLSTM and a 3D-CNN M as reported by Rustowicz et al. [22]. Another example is CNN's multi-temporal approach with channel-attention, which demonstrates an average F1 score of 0.6 for classifying multi-class crops [23]. However, another paper claims that the one-shot classification based on hyperspectral imaging and deep convolutional networks approach is effective [24]. Spatiotemporal transferability is one of the most challenging problems in crop recognition. Recently, a transferable model based on the U-Net++ architecture was proposed [25].

Statistical methods based on Bayesian theory have become widespread for environmental applications and Earth sciences. An example is the deep Bayesian network applied to develop a robust satellite image classifier to prevent adversarial attacks [26]. Moreover,

Bayesian model averaging was used to provide an aggregated probabilistic estimate of soybean crop yield forecast of deep neural networks, such as the 3DCNN (3D Convolutional Neural Network) and ConvLSTM (Convolutional Long Short-Term Memory) [27]. For classifying tree species, Bayesian inference was used, making it possible to achieve an overall classification accuracy of 87% [28]. Recently, Bayesian Deep Image Prior was proposed to downscale soil moisture data from satellite products [29].

### 3. Materials and Methods

#### 3.1. Dataset Description

The dataset used in this work was taken from the ‘Farmpins’ Crop Classification Challenge, which was held on the Zindi challenge platform in 2019 [30]. This dataset consists of time series of satellite images and labeled field masks for South Africa dated in 2017. The area of interest is the proximities of the Orange River—the major agricultural region of South Africa.

The satellite image time-series in the dataset include Sentinel-2 scenes. FarmPin provides satellite images of the entire region across 11 time slices per year covering both summer and winter months. Each Sentinel-2 multi-spectral satellite image has 13 bands, and each band represents the intensities of absorbed radiation by the satellite’s sensor in a specific wavelength range. Bands of the Sentinel-2 L1C images have three different spatial resolutions: 60 m, 20 m, and 10 m for the pixel’s side length.

Labeled field mask data represent field masks for each of nine crops: cotton, dates, grass, lucern, maize, pecan, vineyard, intercrop (mixed vineyard and pecan), and the last one is for vacant fields with no plants on them. The organizers of the challenge claimed that all data were verified by them personally and with the help of drones in 2017.

The initial dataset was split into training and testing datasets, which contain 2497 and 1074 fields, respectively. In this work, only the labeled part was used for training and validating the obtained models; therefore, the final number of fields used in this work was 2497. They were further split into the training and validation sets.

The distribution of fields by classes from the dataset is shown in Table 1.

**Table 1.** Number of fields for each crop after preprocessing of Farmpin dataset.

Cotton	Dates	Grass	Lucern	Maize	Pecan	Vacant	Vineyard	Intercrop
113	2	85	468	251	135	233	789	71

A distinctive feature of the dataset is relatively small areas of the fields. Each field occupies the area of approximately 0.015 km<sup>2</sup> on average. Hence, fields contain around 150 pixels at 10 m resolution, which is the best available spatial resolution of Sentinel-2 images.

Since pixel-wise classification with each pixel being treated independently was primarily used by us, the number of pixels in the Dates class was not enough to train a model. For this reason, the Dates class was removed from the experiments. Additionally, the Intercrop class was omitted as these fields contain both the Vineyard and Pecan classes.

#### 3.2. Satellite Data Preprocessing

The Sentinel 2 L1C satellite images were used as input data for classification. The preprocessing pipeline that has been used in the experiments is described below:

1. Radiometric calibration.  
Radiometric calibration is a procedure used to convert meaningless pixel values (DNs) into physical values such as the intensities of light reflected from all reflecting media and absorbed by the satellite filter. Reflected light from the ground and atmosphere is denoted as Top of the Atmosphere (TOA) reflectance. To derive an image from an uncalibrated image, gain and offset should be applied to raw pixel values (DNs). Typically, information about gain and offset is received from the data provider or obtained from the satellite image metadata.

2. **Atmospheric correction.**  
Since light reflected by plants is worked with, atmospheric correction procedure should be applied, which converts TOA reflectance into surface or BOA reflectance (BOA—bottom of the atmosphere). Both of these procedures, radiometric calibration and atmospheric correction, were performed with the help of the Sentinel Application Platform (SNAP) tool [31]. Radiometric correction is automatically applied to uploaded Sentinel images. Atmospheric correction was carried out using the Sen2Cor plug-in, which is freely distributed by Sentinel.
3. **Spatial resampling.**  
As it was mentioned above, Sentinel images contain bands of different resolutions: 60 m, 20 m, and 10 m per pixel's side length. Bands with the 60 and 20 m resolution were resampled using bilinear interpolation to the resolution of 10 m per pixel's side length.
4. **Band normalization.**  
Band normalization is a standard procedure used to normalize each band of a satellite image. Pixel values (normalized BOA reflectance in this case) were used as an input or baseline of the model. The following formula was used for the calculations:  $\hat{x}_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$ , where  $\hat{x}_{ij}$  is the  $i$ -th pixel in a normalized  $j$ -th band in the image  $x$ ;  $\mu_j$  and  $\sigma_j$  are the mean pixel value of  $j$ -th band and its standard deviation, respectively. This normalization improves the robustness of the trained models.

### 3.3. Feature Generation

Spectral vegetation indices (SVI) were used as additional features in the classification models. SVI quantifies the contribution of vegetation properties and allows reliable spatial and temporal inter-comparisons of terrestrial photosynthetic activity and canopy structural variations. Since different plants have different spectrum reflectance rates, SVI will also differ from one crop to another [32]. Following is a list of the used vegetation indices:

- **Normalized Difference Vegetation Index (NDVI).**  
NDVI is the most well-known vegetation index in agriculture. NDVI is considered to be an indicator of crop health. It has been shown that there is a strong correlation between NDVI and the amount of green vegetative mass [33]. The NDVI was calculated as follows:

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

- **Enhanced Vegetation Index (EVI).**  
EVI is the optimized version of NDVI designed to have a slightly higher sensitivity in the regions with high biomass amounts. In this study, the 2-band version of this index was used [34]. The EVI was calculated as follows:

$$\text{EVI} = 2.5 \frac{\text{NIR} - \text{Red}}{\text{NIR} + 2.4 \text{Red} + 1} . \quad (2)$$

- **Normalized Difference Red Edge Index (NDRE).**  
NDRE is similar to NDVI, except that Red Edge (RE) Band (Band 6) is used instead of Red Band. The NDRE was calculated as follows:

$$\text{NDRE} = \frac{\text{NIR} - \text{RE}}{\text{NIR} + \text{RE}} . \quad (3)$$

- **Modified Soil-Adjusted Vegetation Index (MSAVI).**  
MSAVI is the index used to minimize bare soil effects on the Soil Adjusted Vegetation Index (SAVI) [35]. The modified soil-adjusted vegetation index (MSAVI) is a vegetation index used for areas with large patches of bare soil. These areas are unsuitable for

NDVI due to the small vegetation areas and the absence of chlorophyll. The MSAVI was calculated as follows:

$$\text{MSAVI} = \frac{2\text{NIR} + 1 + \sqrt{(2\text{NIR} + 1)^2 - 8(\text{NIR} - \text{Red})}}{2}. \quad (4)$$

These features were calculated from BOA reflectances before the normalization step of the preprocessing pipeline and were added as additional channels. In total, the input tensor for classification models contained 17 channels.

### 3.4. Resampling Algorithms

Classification of crops via satellite data related to class imbalance data, since in most agricultural regions several major crops and many minor crops are grown [16]. The small area of fields in the study region has an additional impact on the imbalance of classes. In order to overcome high imbalances of classes, resampling techniques were applied to the input data. These techniques are as follows:

- Random Over-Sampling (ROS).  
Random Over-Sampling consists in randomly sampling new objects from the set of available objects with replacement for each class [36].
- Random Under-Sampling (RUS).  
The Random Under-Sampling method is quite similar to ROS. However, instead of sampling new objects, a subset is formed from the existing one by randomly choosing objects with equal probability for every sample [37].
- Synthetic Minority Over-Sampling Technique (SMOTE).  
SMOTE [38] is a more advanced technique for dataset resampling, where, in contrast to ROS and RUS, only new samples are created. For a randomly picked sample from the dataset  $x_{base}$ , a set of  $k$ -nearest neighbors is considered. Then, a corresponding sample  $x_{corr}$  is chosen from the set of these neighbors' objects. The new sample is generated according to  $x_{new} = x_{base} + \lambda (x_{corr} - x_{base})$ , where  $\lambda \in [0, 1]$  and is sampled using uniform distribution.

Resampling techniques to overcome unbalanced classes used with ML models, such as ROS, RUS and SMOTE are not suitable for convolutional neural networks. Instead of resampling techniques, it is a common practice to assign the weights for each class and use them in the loss function [39]. It allows to reduce the effect of majority class gradient direction dominance in the backpropagation process during training.

The weighted loss is calculated according to the formula:

$$\mathcal{L}_w(D) = \frac{1}{|D|} \sum_{(x,y) \in D} w_y \mathcal{L}(x, y), \quad (5)$$

where  $\mathcal{L}$  is a certain loss function,  $D$  is a dataset,  $(x, y)$  is a sample-label pair and  $w_y$  is the weight for class  $y$ .

### 3.5. Data Augmentation

A data augmentation technique was used to artificially increase the dataset by transforming the initial dataset samples to prevent model overfitting and even the imbalances between classes. Augmentation methods were applied for classification tasks using multi-spectral images [40,41]. The Python library Albumentations was used to build the augmentation pipeline [42]. In the framework here, 100 by 100 pixel images we used. The following augmentation pipeline was used:

1. Pad the image with reflection mode by 50 pixels on each side;
2. Rotate the image by a random angle in the  $(-60, 60)$  degree range;
3. Apply affine transformation with random parameters in the  $(0.75, 1)$  range;
4. Shift image by a random number of pixels in the  $(-20, 20)$  range along both axes;

### 5. Crop in the center down to the size of 100 by 100 pixels.

In all image transforms, bilinear interpolation was used; in label transforms, 0-order transformations were used. Both satellite images and labels were transformed using this pipeline. Output augmented images were fed to the neural network after that.

### 3.6. Classical Pixel Aggregation Methods

We used a pixel-wise approach to train the model, with probabilities  $p(k|x)$  of a certain pixel being represented by the feature vector  $x$  to the class  $k$ . However, the results of field classification may contain more valuable data for farmers and decision-makers than those generated by pixel-based classification; for this reason, we focused on this type of crop classification. As the classification method has a certain degree of misclassification, a number of pixels within fields were assigned to the wrong classes. At the same time, there is prior information that all pixels from a particular field belong to one class; therefore, we decided to remove misclassified pixels by assigning all pixels within one field to a single class. The classical methods applied for this purpose are average voting and majority voting.

#### 3.6.1. Majority Voting

In machine learning, majority voting is a primary type of aggregation, which enables algorithms to make decisions about the best classifications. For our case, it is formulated as follows: for each pixel  $x_i$ , we deduce its prediction  $c_i$  by taking a class with maximum probability and then choosing for the whole field the class  $c$  with the maximum number of classified pixels. Here,  $x_i$ —features of the  $i$ -th pixel of the field,  $p(\text{class} = k | x_i)$ —probability of the  $i$ -th pixel being classified to class  $k$ , and  $\#c_n$  is the number of pixels classified to class  $c_n$ . Formulae for this rule:

$$c_i = \arg \max_k p(\text{class} = k | x_i), \quad (6)$$

$$c = \arg \max_n \#c_n \quad (7)$$

The main disadvantage of this aggregation rule is that it does not take into account the confidence of prediction for each pixel. In a situation where the majority class has fewer pixels but with a more confident prediction than the minority class, the less confident class will be chosen.

#### 3.6.2. Average Voting

Averaging is another popular machine learning strategy. It takes the mean predicted probability  $\bar{p}(\text{class} = k)$  of each pixel's class  $c$  among all crops in a field  $F$  and uses that as the final class for that field:

$$\bar{p}(\text{class} = k) = \frac{1}{|F|} \sum_{i \in F} p(\text{class} = k | x_i), \quad (8)$$

$$c = \arg \max_k \bar{p}(\text{class} = k). \quad (9)$$

Average voting is a useful technique; however, the result of this aggregation method can be misleading. Usually, each pixel is treated as a new source of information about the whole field, especially when one has a field with quite a few pixels. Each new pixel that predicts some class with positive probability (greater than 0.5) should increase the confidence of the overall prediction of that class. However, that does not necessarily happen when using average aggregation. For example, consecutive pixels with 'positive' (higher than 0.5) probabilities—0.6, 0.7, 0.8—will obtain 0.7 as the overall probability of some class (here, the pixel with probability 0.6 canceled out the one with a high probability of 0.8), although they all are 'positive' and final confidence should be higher than 0.8.

### 3.7. Bayesian Aggregation

The natural way of incorporating prior information about a pixel's class into a decision is to use Bayes' theorem. Let  $\mathbb{F}$  be the set of pixel features from some field:  $x_i \in \mathbb{F}$ . Summing up the log-odds of predictions from every pixel (10), the final decision is obtained using (11):

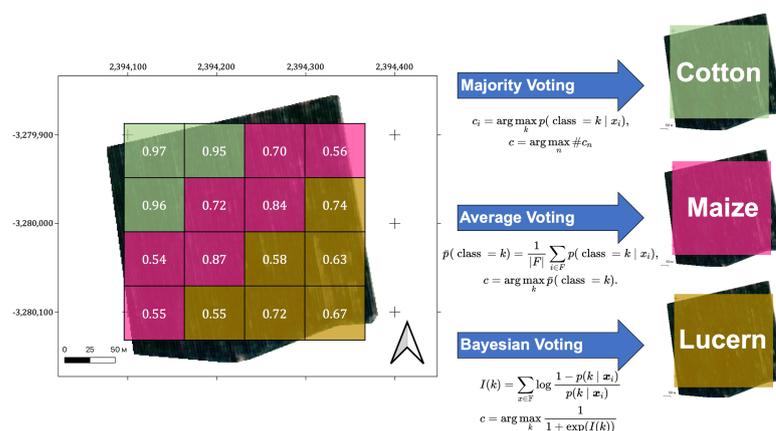
$$I(k) = \sum_{x \in \mathbb{F}} \log \frac{1 - p(k | x_i)}{p(k | x_i)}, \quad (10)$$

$$c = \arg \max_k \frac{1}{1 + \exp(I(k))}. \quad (11)$$

Before applying Bayesian aggregation, it may be useful to perform smoothing of raw predictions in the following way:

$$\hat{p}_\alpha(k | x_i) = \alpha p(k | x_i) + \frac{1 - \alpha}{N - 1} (1 - p(k | x_i)), \quad (12)$$

where  $\alpha \in (0, 1)$  is a smoothing factor and  $N$  is the number of classes for classification. The choice of factor  $\alpha$  depends on the methods used. A grid search was conducted here, which revealed that for  $N = 7$  as in the present case, values  $\alpha \in (0.3, 0.4)$  perform best. Figure 1 presents the process of pixel aggregation.



**Figure 1.** Theoretical visualization of pixel aggregation for the classification model output. Average voting, majority voting and Bayesian voting produce different classes for fields with some pixels assigned to the wrong class.

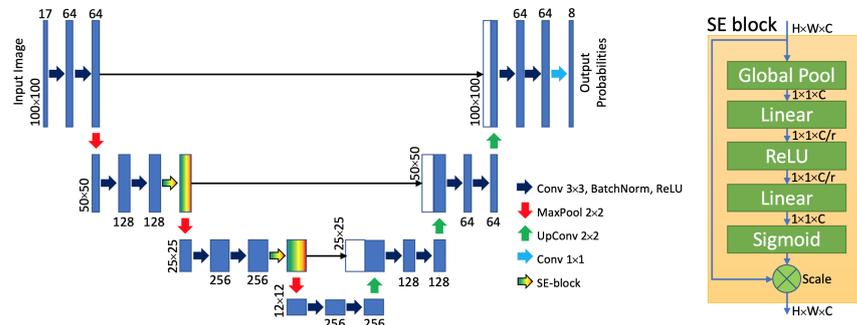
### 3.8. Crop Classification Pipeline

Using preprocessed bands and generated features, ML models and U-Net were trained and pixels were classified into classes according to the field dataset. First, satellite data were preprocessed and SVIs were computed as additional features. Second, the initial dataset with crop labels was divided into the training and test sets. After that, ML models and U-Net were trained using different resampling techniques and pixel-wise classified results were obtained. The following machine learning and deep learning methods were applied:

- k-nearest neighbors (kNN) [43];
- Random Forest (RF) [44,45];
- Gradient Boosting Decision Trees (GB) [46];
- U-Net architecture with SE-blocks [47].

The architecture of the U-Net and SE-blocks network is shown in the Figure 2. Finally, pixel aggregation techniques were used to forecast crop types on a field-scale level. To evaluate the performance of the ML pipeline, overall accuracy and Macro F1-score were

used [48]. Hyperparameters and other technical details of the applied machine learning models are presented in Appendix A.



**Figure 2.** Architecture of U-Net with SE-blocks used for crop classification.

#### 4. Results and Discussion

In this section, we present the results of our experiments on crop classification, as well as the comparison of various resampling techniques and pixel aggregation methods.

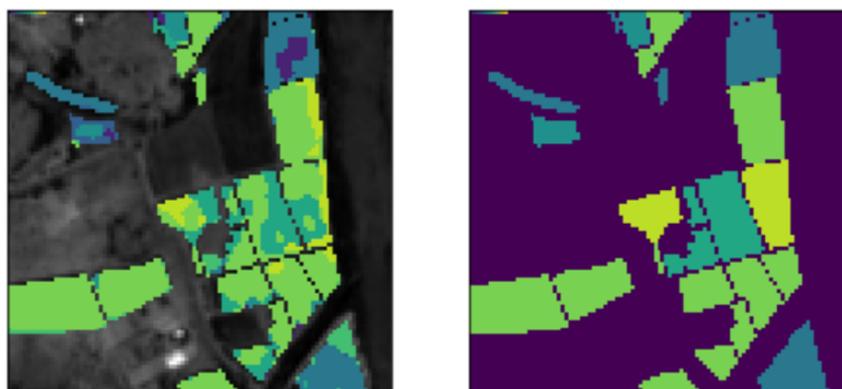
Table 2 shows the main results of pixel-wise crop classification before applying the aggregation procedure. It provides the assessment of ML models and U-Net neural network performance using overall accuracy and F1 score. The table also compares the results of different resampling techniques: ROS, RUS, SMOTE and weighting.

**Table 2.** Comparison of resampling techniques and pixel-wise classification results for ML models and U-net before pixel aggregation. The table does not contain metrics for resampling methods incompatible with some algorithms. For example, ROS, RUS, and SMOTE methods are not applicable for CNN, and weighting is not for kNN.

Classifier	ROS		RUS		SMOTE		Weighting	
	OA	Macro F1	OA	Macro F1	OA	Macro F1	OA	Macro F1
kNN	53.0	0.39	43.5	0.33	53.3	0.40	—	—
RF	62.7	0.46	50.3	0.37	62.4	0.46	62.5	0.46
GB	68.5	0.51	55.3	0.42	65.7	0.50	68.7	0.50
U-Net+SE	—	—	—	—	—	—	70.1	0.57

Table 2 demonstrates the differences in the performances between ML models and U-Net. The U-Net neural network with SE-blocks (U-Net+SE) outperformed all other methods, namely, GB, kNN classifier and RF. Its OA was as high as 70.1 and Macro F1 score as high as 0.57. When classical machine learning models were compared, GB showed the best result (OA 68.7, Macro F1 0.50). The comparison of resampling techniques demonstrated that all the techniques used give more or less similar results in terms of the Macro F1 score. However, the choice of resampling technique greatly affected the OA metric. Weighting resampling outperforms the RUS technique by more than 10% in the case of gradient boosting (GB). Generally, Weighting resampling appears to be a better choice than ROS, RUS and SMOTE to tackle the unbalanced classes problem.

Figure 3 presents the result of crop classification using the U-Net neural network. Different classes of crops on the map are marked with different colors. As we can see on the map, there are fields containing almost equal numbers of pixels assigned to different classes. Such cases pose challenges in field-wise classification, as it is unclear how to make a final decision about the crop class. To address this issue, we tested various aggregation techniques.



**Figure 3.** Classification map produced by U-Net+SE setup (**left**) and correct crop labels (**right**).

Table 3 presents the comparison of different aggregation techniques performance towards field-based crop classification. The effect of aggregation on the results is promising: we applied three different aggregation approaches, and all of them led to a marked improvement in crop recognition results at field scale. Bayesian aggregation showed the best OA among the tested methods. The U-Net neural network shows similar OA metrics for both pixel-wise and field-wise crop classifications, probably because U-net considers particular contextual information.

**Table 3.** Comparison of different aggregation techniques and resampling techniques for field-based crop classification by ML models and U-Net. The table does not contain metrics for resampling methods incompatible with some algorithms. For example, ROS, RUS, and SMOTE methods are not applicable for CNN, and weighting is not for kNN. Bold text indicates the best overall accuracy for each of the resampling methods. For all methods, the Bayesian strategy of pixel aggregation improved overall accuracy.

		Overall Accuracy (%)			
	Aggregation	kNN	RF	GB	U-Net+SE
ROS	Bayesian	72.66	73.61	77.44	—
	Averaging	72.28	73.42	77.25	—
	Majority	72.08	72.08	76.67	—
RUS	Bayesian	61.41	60.11	70.44	—
	Averaging	61.41	59.92	70.08	—
	Majority	58.62	58.48	69.69	—
SMOTE	Bayesian	72.85	73.23	77.44	—
	Averaging	73.04	73.04	77.44	—
	Majority	72.66	71.70	75.72	—
Weighting	Bayesian	—	72.85	77.44	71.58
	Majority	—	72.47	77.06	71.34
	Averaging	—	71.13	75.91	71.22

According to our results, RF, kNN and GB performed sufficiently well in the case of crop classification based on a single satellite image. It also appears that ROS and weighting techniques produce the best overall accuracy and Macro F1 score among the tested methods. To the best of our knowledge, this is the first study where the Bayesian aggregation approach was successfully applied to improve crop classification at the field scale.

## 5. Conclusions

We compared the performances of classical machine learning methods and the U-Net neural network for the crop classification task in South Africa and achieved reasonable performance using aggregation strategy. We assessed and compared the performance of

different classical aggregation strategies and suggested a new one based on Bayes theorem, which has never been tested in applied tasks. Bayesian aggregation outperformed other aggregation strategies—namely, majority voting by 1.5% and averaging approach by 0.6%. Consequently, crop classification with Bayesian aggregation is a promising approach for industrial use in agriculture. Moreover, it could be used in other geospatial applications for classification goals, such as forest taxation, land use and land types recognition and others.

**Author Contributions:** Conceptualization, I.M., A.P., R.J., M.P. and I.O.; Funding acquisition, M.K.; Methodology, I.M., M.G., A.P. and M.P.; Project administration, M.P. and I.O.; Resources, M.K. and M.P.; Software, I.M. and M.G.; Supervision, A.P.; Visualization, M.G.; Writing—original draft, I.M.; Writing—review and editing, M.G., A.P., R.J., M.K. and I.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No. 075-15-2020-801).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Input data used for research are freely available from the Zindi Platform—<https://zindi.africa/competitions/farm-pin-crop-detection-challenge/> (accessed on 1 September 2022).

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

## Abbreviations

The following abbreviations are used in this manuscript:

OA	Overall Accuracy
ROS	Random Over-Sampling
RUS	Random Under-Sampling
SMOTE	Synthetic Minority Over-Sampling Technique
NDVI	Normalized Difference Vegetation Index
EVI	Enhanced Vegetation Index
NDRE	Normalized Difference Red Edge Index
MSAVI	Modified Soil-Adjusted Vegetation Index
kNN	k-Nearest Neighbors
RF	Random Forest
GB	Gradient Boosting Decision Trees
DL	Deep Learning
ML	Machine Learning
NNs	Neural Networks

## Appendix A. Technical Details

### General information

All parts of the project were implemented using Python 3.7 programming language and Jupyter Notebook. For the implementation of the neural networks, PyTorch 1.5.0 package was used. Models were trained using an NVidia GTX 970 GPU. Models such as k-nearest neighbors and Random Forest were imported from the Sci-Kit Learn Python package. For the gradient boosting machine, the XGBoost library was used. For all image transforms used in the data augmentation part of the data preparation pipeline, functions implemented in the Albumentations Python package were imported.

### SMOTE

Number of neighbors: 8

k-Nearest Neighbors

Number of neighbors: 8

Points are weighted by the inverse of their distance

Random Forest:

Number of estimators: 200  
 Minimal number of samples in the leaf: 10  
 Maximum depth of tree: 15  
 Class weights are inversely proportional to class frequencies  
 Gradient Boosting Decision Trees:  
 Number of estimators: 250  
 Maximum depth of tree: 10  
 Subsample rate: 0.5  
 Learning rate: 0.05  
 Class weights are inversely proportional to class frequencies  
 U-Net with Squeeze and Excitation Blocks  
 Training (2 heads):  
 Optimization algorithm: Adam  
 Learning rate: 0.01  
 Batch size: 8  
 Shuffling when training: True  
 Learning rates were decreased by the factor of 0.977 every epoch  
 Number of epochs: 350  
 Fine-tuning:  
 Learning rate: 0.000002  
 Learning rates were decreased by the factor of 0.984 every epoch  
 Number of epochs: 100

## References

1. Ma, L.; Li, M.; Ma, X.; Cheng, L.; Du, P.; Liu, Y. A review of supervised object-based land-cover image classification. *Isprs J. Photogramm. Remote Sens.* **2017**, *130*, 277–293. [[CrossRef](#)]
2. Schultz, B.; Immitzer, M.; Formaggio, A.; Sanches, I.; Luiz, A.; Atzberger, C. Self-guided segmentation and classification of multi-temporal Landsat 8 images for crop type mapping in southeastern Brazil. *Remote Sens.* **2015**, *7*, 14482–14508. [[CrossRef](#)]
3. Vuolo, F.; Neuwirth, M.; Immitzer, M.; Atzberger, C.; Ng, W.T. How much does multi-temporal Sentinel-2 data improve crop type classification? *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 122–130. [[CrossRef](#)]
4. Shukla, G.; Garg, R.D.; Srivastava, H.S.; Garg, P.K. Performance analysis of different predictive models for crop classification across an aridic to ustic area of Indian states. *Geocarto Int.* **2018**, *33*, 240–259. [[CrossRef](#)]
5. Arafat, S.M.; Aboelghar, M.A.; Ahmed, E.F. Crop discrimination using field hyper spectral remotely sensed data. *Adv. Remote Sens.* **2013**, *2*, 63. [[CrossRef](#)]
6. Sonobe, R.; Yamaya, Y.; Tani, H.; Wang, X.; Kobayashi, N.; Mochizuki, K.i. Assessing the suitability of data from Sentinel-1A and 2A for crop classification. *Giscience Remote Sens.* **2017**, *54*, 918–938. [[CrossRef](#)]
7. Lussem, U.; Hütt, C.; Waldhoff, G. Combined Analysis of SENTINEL-1 and Rapideye Data for Improved Crop Type Classification: AN Early Season Approach for Rapeseed and Cereals. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 959–963. [[CrossRef](#)]
8. Arvor, D.; Dubreuil, V.; Simões, M.; Bégué, A. Mapping and spatial analysis of the soybean agricultural frontier in Mato Grosso, Brazil, using remote sensing data. *GeoJournal* **2013**, *78*, 833–850. [[CrossRef](#)]
9. Chen, Y.; Lu, D.; Moran, E.; Batistella, M.; Dutra, L.V.; Sanches, I.D.; da Silva, R.F.B.; Huang, J.; Luiz, A.J.B.; de Oliveira, M.A.F. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *69*, 133–147. [[CrossRef](#)]
10. Zheng, B.; Myint, S.W.; Thenkabail, P.S.; Aggarwal, R.M. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 103–112. [[CrossRef](#)]
11. Kang, J.; Zhang, H.; Yang, H.; Zhang, L. Support Vector Machine Classification of Crop Lands Using Sentinel-2 Imagery. In Proceedings of the 2018 7th International Conference on Agro-geoinformatics (Agro-geoinformatics), Hangzhou, China, 6–9 August 2018; pp. 1–6.
12. Gilbertson, J.K.; Kemp, J.; Van Nekerck, A. Effect of pan-sharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques. *Comput. Electron. Agric.* **2017**, *134*, 151–159. [[CrossRef](#)]
13. Asgarian, A.; Soffianian, A.; Pourmanafi, S. Crop type mapping in a highly fragmented and heterogeneous agricultural landscape: A case of central Iran using multi-temporal Landsat 8 imagery. *Comput. Electron. Agric.* **2016**, *127*, 531–540. [[CrossRef](#)]
14. Kumar, P.; Prasad, R.; Choudhary, A.; Mishra, V.N.; Gupta, D.K.; Srivastava, P.K. A statistical significance of differences in classification accuracy of crop types using different classification algorithms. *Geocarto Int.* **2017**, *32*, 206–224. [[CrossRef](#)]
15. Asam, S.; Gessner, U.; Almengor González, R.; Wenzl, M.; Kriese, J.; Kuenzer, C. Mapping Crop Types of Germany by Combining Temporal Statistical Metrics of Sentinel-1 and Sentinel-2 Time Series with LPIS Data. *Remote Sens.* **2022**, *14*, 2981. [[CrossRef](#)]

16. Waldner, F.; Chen, Y.; Lawes, R.; Hochman, Z. Needle in a haystack: Mapping rare and infrequent crops using satellite imagery and data balancing methods. *Remote Sens. Environ.* **2019**, *233*, 111375. [CrossRef]
17. Özdemir, A.; Polat, K.; Alhudhaif, A. Classification of imbalanced hyperspectral images using SMOTE-based deep learning methods. *Expert Syst. Appl.* **2021**, *178*, 114986. [CrossRef]
18. Maus, V.; Câmara, G.; Cartaxo, R.; Sanchez, A.; Ramos, F.M.; De Queiroz, G.R. A time-weighted dynamic time warping method for land-use and land-cover mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3729–3739. [CrossRef]
19. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [CrossRef]
20. Ji, S.; Zhang, C.; Xu, A.; Shi, Y.; Duan, Y. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sens.* **2018**, *10*, 75. [CrossRef]
21. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [CrossRef]
22. M Rustowicz, R.; Cheong, R.; Wang, L.; Ermon, S.; Burke, M.; Lobell, D. Semantic Segmentation of Crop Type in Africa: A Novel Dataset and Analysis of Deep Learning Methods. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 15–20 June 2019; pp. 75–82.
23. Tang, P.; Du, P.; Xia, J.; Zhang, P.; Zhang, W. Channel attention-based temporal convolutional network for satellite image time series classification. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
24. Meng, S.; Wang, X.; Hu, X.; Luo, C.; Zhong, Y. Deep learning-based crop mapping in the cloudy season using one-shot hyperspectral satellite imagery. *Comput. Electron. Agric.* **2021**, *186*, 106188. [CrossRef]
25. Wang, L.; Wang, J.; Zhang, X.; Wang, L.; Qin, F. Deep segmentation and classification of complex crops using multi-feature satellite imagery. *Comput. Electron. Agric.* **2022**, *200*, 107249. [CrossRef]
26. Pang, Y.; Cheng, S.; Hu, J.; Liu, Y. Robust Satellite Image Classification with Bayesian Deep Learning. In Proceedings of the 2022 Integrated Communication, Navigation and Surveillance Conference (ICNS), Dulles, VA, USA, 5–7 April 2022; pp. 1–8.
27. Abbaszadeh, P.; Gavahi, K.; Alipour, A.; Deb, P.; Moradkhani, H. Bayesian multi-modeling of deep neural nets for probabilistic crop yield prediction. *Agric. For. Meteorol.* **2022**, *314*, 108773. [CrossRef]
28. Axelsson, A.; Lindberg, E.; Reese, H.; Olsson, H. Tree species classification using Sentinel-2 imagery and Bayesian inference. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *100*, 102318. [CrossRef]
29. Fang, Y.; Xu, L.; Chen, Y.; Zhou, W.; Wong, A.; Clausi, D.A. A Bayesian Deep Image Prior Downscaling Approach for High-Resolution Soil Moisture Estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4571–4582. [CrossRef]
30. Organizers, F.C. Farm Pin Crop Detection Challenge. Available online: <https://zindi.africa/competitions/farm-pin-crop-detection-challenge> (accessed on 24 February 2020).
31. Louis, J.; Debaecker, V.; Pflug, B.; Main-Knorn, M.; Bieniarz, J.; Mueller-Wilm, U.; Cadau, E.; Gascon, F. Sentinel-2 Sen2Cor: L2A processor for users. In Proceedings of the Living Planet Symposium 2016, Prague, Czech Republic, 9–13 May 2016; pp. 1–8.
32. Kyllö, K. *NASA Funded Research on Agricultural Remote Sensing*; Department of Space Studies, University of North Dakota: Grand Forks, ND, USA, 2003.
33. Carlson, T.N.; Ripley, D.A. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens. Environ.* **1997**, *62*, 241–252. [CrossRef]
34. Jiang, Z.; Huete, A.R.; Didan, K.; Miura, T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* **2008**, *112*, 3833–3845. [CrossRef]
35. website editor, E. Modified Soil-Adjusted Vegetation Index. Available online: <https://eos.com/agriculture/msavi/> (accessed on 13 May 2020).
36. Chawla, N.V. Data mining for imbalanced datasets: An overview. In *Data Mining and Knowledge Discovery Handbook*; Springer: Boston, MA, USA, 2009; pp. 875–886.
37. Batista, G.E.; Prati, R.C.; Monard, M.C. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explor. Newsletter* **2004**, *6*, 20–29. [CrossRef]
38. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
39. Bischke, B.; Helber, P.; Borth, D.; Dengel, A. Segmentation of imbalanced classes in satellite imagery using adaptive uncertainty weighted class loss. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 6191–6194.
40. Illarionova, S.; Nesteruk, S.; Shadrin, D.; Ignatiev, V.; Pukalchik, M.; Oseledets, I. MixChannel: Advanced augmentation for multispectral satellite images. *Remote Sens.* **2021**, *13*, 2181. [CrossRef]
41. Yan, S.; Yao, X.; Zhu, D.; Liu, D.; Zhang, L.; Yu, G.; Gao, B.; Yang, J.; Yun, W. Large-scale crop mapping from multi-source optical satellite imageries using machine learning with discrete grids. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102485. [CrossRef]
42. Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albumentations: Fast and Flexible Image Augmentations. *Information* **2020**, *11*, 125. [CrossRef]
43. Cunningham, P.; Delany, S.J. k-Nearest neighbour classifiers-A Tutorial. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–25. [CrossRef]
44. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
45. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [CrossRef]

46. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
47. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241.
48. Grandini, M.; Bagli, E.; Visani, G. Metrics for multi-class classification: An overview. *arXiv* **2020**, arXiv:2008.05756 .