**RESEARCH**

# Impact of segmentation algorithms on multisensor LULC classification in a semiarid Mediterranean area

Carmen Valdivieso-Ros[1,2] · Francisco Alonso-Sarría[1,2] · Francisco Gomariz-Castillo[1,2]

## Abstract

Geographic object-based image analysis methods usually provide better results than pixel-based methods for classifying land use and land cover from high and medium resolution satellite imagery. This study compares the results of Random Forest (RF) and Multilayer Perceptron (MLP) when used to classify the segments obtained on an RGB+NIR Sentinel-2 image using three different segmentation algorithms, Multiresolution (MR), Region Growing (RG), and Mean-Shift (MS). The hyperparameters of these algorithms were optimised minimising the intra-object heterogeneity and maximizing the inter-object heterogeneity, integrating them in an optimization loop. Geometric and two different centrality and dispersion statistics were computed from some Sentinel-1, Sentinel-2 and LiDAR variables over the segments, and used as features to classify the datasets. The highest segment cross-validation accuracies were obtained with RF using MR segments: 0.9048 (k=0.8905), while the highest accuracies calculated with test pixels were obtained with MLP using MR segments: 0.9447 (k=0.9303), both with the mean and standard deviation of the feature set. Although the overall accuracy is quite high, there are problems with some classes in the confusion matrix and, significant misclassification appear when a qualitative analysis of the final maps is performed, indicating that the accuracy metrics may be overestimated and that a qualitative analysis of the results may also be necessary.

**Keywords** GEOBIA · Random forest · Multilayer perceptron · Segmentation · Multisensor data

## Introduction

Geographic object-based image analysis (GEOBIA) methods tend to outperform pixel-based methods when applied to land cover and land use classification using high-resolution satellite imagery (Gao and Mas 2008; Whiteside et al. 2011; Jozdani et al 2019; Šiljeg et al. 2022). This is because they are able to identify spatial objects above the minimum image unit, the pixel, which improves pattern recognition

✉ Francisco Alonso-Sarría
  alonsarp@um.es

  Carmen Valdivieso-Ros
  mcarmen.valdivieso@um.es

  Francisco Gomariz-Castillo
  fjgomariz@um.es

[1] Departamento de Geografía, Universidad de Murcia, Campus La Merced, Murcia 30001, Spain

[2] Instituto Universitario del Agua y del Medio Ambiente, Universidad de Murcia, Campus de Espinardo, Murcia 30100, Spain

and reduces the computational cost of classification algorithms (Blaschke et al. 2014). The analysis starts with image segmentation, i.e. the separation of the image into homogeneous regions or objects. Then, relevant features of each object (such as shape, size, texture or colour, among others) are extracted to classify these objects (Lang 2008).

Initially, this type of analysis was oriented towards high resolution imagery. Although GEOBIA is mainly used with high resolution imagery, the widespread use of medium resolution imagery from satellites such as Landsat or Sentinel, and the synergy of optical sensor data with data from other sources such as SAR or Lidar, has broadened the application of GEOBIA. Ruiz et al. (2021) successfully applied multi-source data from SAR and optical Sentinel sensors and GEOBIA techniques to classify wetland vegetation classes. Other data sources have also been explored to perform a large area land cover classification, e.g. Maxwell et al. (2019) use an orthophoto in combination with a DEM and other ancillary data such as population or road density. Gbodjo et al. (2020) proposed an architecture of a neural network using multi-source and multi-temporal data to a GEOBIA LULC mapping in several locations, with disparate results and Wu

et al. (2016) explored the application of a multi-source approach integrating LiDAR and optical data to perform an object-oriented hierarchical classification of urban land cover.

However, image segmentation is a complicated procedure with many difficulties to be solved. The neighbouring pixels of an image have to be grouped according to unknown spatial, spectral and/or contextual criteria, which have to be defined a priori (Räsänen et al. 2013). The scale problem falls under this aspect, since the decision of the optimal parameters must take into account that the objects may involve different scales (Georganos et al 2018). The quality of the subsequent classification results will depend on the choice of these criteria (Blaschke et al. 2008). Thus, the best segmentation will be the one that minimises the internal spectral heterogeneity of the segments and maximises the external between them. Therefore, the first problem to be faced when performing this type of analysis is to find the best combination of parameters to obtain the best possible segmentation according to the internal and external heterogeneity. However, how to find this combination is still considered one of the most difficult problems by the scientific community (Johnson and Ma 2020).

Image segmentation techniques may be divided into two basic approaches (Anjna and Kaur 2017): (i) those based on discontinuity detection, such as the change in the intensity of the values of a region of the image, which implies the identification of edges; (ii) and those based on similarity detection, called region-based methods, such as region-growing or clustering techniques. More recently, deep learning techniques such as neural network architectures like CNN or U-Net have been increasingly used for segmentation (Ma et al. 2019), although it has been found that they are better suited to high resolution than the medium resolution of Landast or Sentinel-2 (Liu et al. 2019; Wurm et al. 2019).

In region-based segmentation methods, the image is divided into homogeneous regions that resemble objects or regions of interest (e.g., buildings, crops, water bodies, etc.). The segmentation process involves several parameters that need to be optimised, such as the scale (i.e., the size of these objects or regions) or the spectral similarity threshold, or the shape and size of the segments. Thus, some of the problems of any parameter optimisation task appear in image segmentation. Another issue is the choice of a segmentation algorithm (Hanbury 2009; Hossain and Chen 2019), as they are computationally intensive due to the need to compute the similarity between pixels and to optimise the parameters.

Quantifying the quality of a segmentation is another problem. There are not many methods to evaluate a segmentation, both visually and subjectively (qualitatively) as well as objectively (quantitatively). The most common problems in image segmentation are oversegmentation and undersegmentation. The latter may contain several classes in a single segment, so it is usually difficult to assign a segment to a single class

in the classification process. On the other hand, an oversegmented image does not present this problem. Therefore, one of the ways to objectively evaluate a segmentation is using the final metrics of the classification (Räsänen et al. 2013).

In addition, most of the work in GEOBIA has been developed using black box models in expensive proprietary software that can't be reproduced and analysed without it (Grippa et al. 2017). Shepherd et al. (2019) developed a free software package that uses k-means to do an initial clustering of the pixels and then merges smaller sets until a minimum cluster size is reached. However, it requires advanced programming skills. Grippa et al. (2017) have developed several modules for the GRASS-GIS software to perform segmentation, optimise the selection of segmentation parameters, and obtain segmentation metrics. However, there is still a notable lack of free, powerful, robust and user-friendly software that includes segmentation methods among its tools. This makes it difficult to apply GEOBIA to images of large areas, as well as to multi-sensor or multi-temporal data, due to the large amount of data to be processed. For the same reasons, the quantitative comparison between different segmentation algorithms is compromised and difficult to perform. All these problems are still the subject of research by the scientific community (Johnson and Ma 2020).

A different but no less important decision is which features to extract from the segments. Centrality and dispersion statistics and others related to shape or size, as in Cnovas-Garca and Alonso-Sarra (2015), are usually extracted from segments. However, the choice between mean or median, or between standard deviation or range, can affect the accuracy results, depending on the segments obtained after the segmentation process. Mean and standard deviation seem to be the most sensible option when the intra-class variance of some of the classes in the classification scheme or the segmentation seems to fit the image objects reasonably well. However, if the data distribution is not normal, median and range would be a more robust choice.

Machine Learning (ML) algorithms such as Random Forest (RF) have shown satisfactory performance for multi-sensor classification in small agricultural areas (Masiza et al. 2020), urban areas (Dobrinić et al 2020) or forests (Luca et al. 2022). Amoakoh et al. (2021) tested multi-sensor classification using RF aiming to accurately monitor a tropical peatland, and Chen et al. (2022) used RF to effectively map forest above-ground biomass in heterogeneous mountainous regions. Cánovas-García and Alonso-Sarría (2015) compared the use of five classification algorithms applied to a high-resolution image segmentation; RF and Support Vector Machines (SVM) produced the most accurate results. However, when spatial or temporal patterns in land use and cover are important, it may be difficult for traditional ML algorithms to extract some contextual features from images (Reichstein et al 2019). For this reason, the use of Artificial

Neural Network (ANN) architectures in LULC classification for image analysis has made great progress. However, the superiority of the performance of Deep Learning (DL) methods over ML methods remains unclear (Wang et al 2022). There are still several unresolved aspects, such as the use of multi-source data (Reichstein et al 2019; Qin and Liu 2022) or the quality and quantity of the training data set (Qin and Liu 2022). In addition, DL has a challenge related to the size of the training data. Applying good performance models with their training and test data to other datasets does not always achieve the same level of success (Qin and Liu 2022), possibly due to an inadvertent form of overtraining. Since algorithms such as Random Forest are robust to small training data sets, ML approaches are still widely used. However, some techniques based on neural networks such as the Multi-Layer Perceptron are being used for comparison with traditional ML algorithms, such as RF or SVM (Valdivieso-Ros et al. 2023).

The aim of this research is to compare the results of different open source segmentation algorithms on Sentinel-2 imagery. Open source software is usually based on libraries that are easier to integrate with optimization routines than proprietary software. The selection of the best segmentation hyperparameters is done by minimising the intra-object heterogeneity and maximising inter-object heterogeneity. The resulting set of objects is classified using both Random Forest and a Multilayer Perceptron, and the final test of both classifications is performed using a set of test pixels outside of the training objects used to calibrate the models. In this way the quality of the segmentation is not determined by the properties of the objects, as they were the objective function in the segmentation optimization, but by the ability of the segmentation to produce an accurate LULC map when its segments are used to calibrate the classification model. We also want to analyse the effect of using two combinations of geometric information with statistics of centrality and dispersion of the predictor variables of each segment on the accuracy of the classification.

## Methods

### Study area

The study area is the Mar Menor basin, located on the south-east coast of Spain, on the Iberian Peninsula. Administratively, it belongs to the Autonomous Community of the Region of Murcia (Fig. 1).

The basin is characterised by a slope of less than 10% from northwest to southwest, draining into a coastal lagoon,
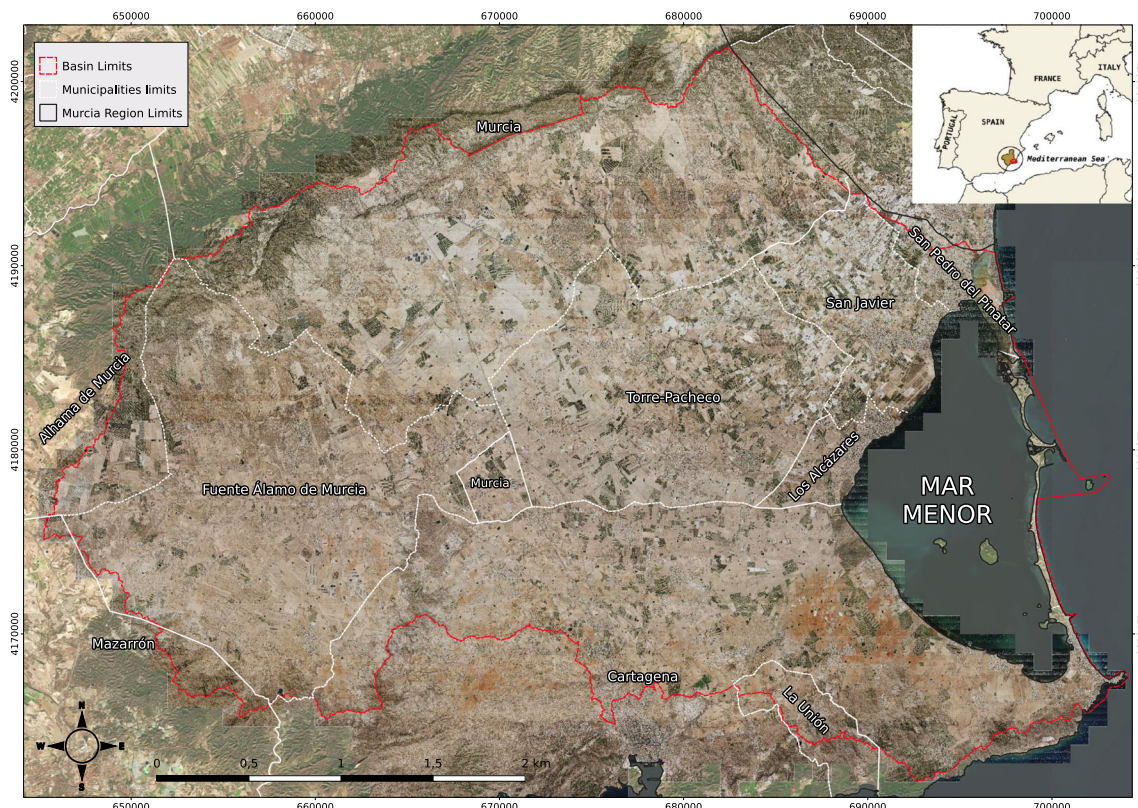


**Fig. 1** Study Area: Mar Menor Basin

the *Mar Menor*. The climate is semi-arid Mediterranean, with high aridity and irregular rainfall. The average annual rainfall is less than 300-350 mm, depending on the proximity to the coast. In addition, the high temporal and spatial variability of rainfall makes alternating droughts and extreme floods common. Temperatures are warm throughout the year, with an average between $16^{o}$C and $18^{o}$C and an average annual maximum that can exceed $42^{o}$C.

The Mar Menor is the largest coastal saltwater lagoon in the western Mediterranean. It is partially enclosed by a sand barrier 22 km long and up to 1.2 km wide, known as *La Manga del Mar Menor*. This lagoon and part of its surroundings are recognised by the most important European protection figures for their high singularity, ecological value and great biodiversity; however, the development of the main uses in the basin, agricultural and urban, has been affecting the lagoon's marine ecosystem for decades (Martínez et al. 2007; Giménez-Casalduero et al. 2020).

Traditional rainfed agriculture has gradually been abandoned or transformed into irrigated agriculture, which is more profitable given the rainfall patterns of the area. According to the regional agricultural statistics (CARM 2023), traditional rainfed agriculture currently accounts for less than 4,500 ha, mainly almond trees, olive groves and some vineyards. Thus, irrigated crops of different classes reached almost 38,000 ha in 2018 and greenhouses about 1,500 ha of surface, although this amount does not include other types of cover such as nets or meshes to prevent damage to the crop by birds and insects or meteorological events such as hail. These nets make a significant difference to the spectral signature and must be taken into account when classifying land cover.

The second most important cover in this area is urban (CARM 2023). There are large urbanised areas along the lagoon's coastline, whose population growth during the holiday season is difficult to quantify.

## Workflow

Figure 2 shows the work flow designed for this study. The first stage is the pre-segmentation of an RGB-Nir image of the study area into sub-regions, followed by the optimisation of the parameters for each segmentation algorithm by each different sub-region.

In parallel, the four different date images from Sentinel-2 and Sentinel-1 are pre-processed and the indices and texture variables, as well as the variables derived from the LiDAR data, are calculated. To build the final dataset, the centrality and dispersion statistics of the variables are calculated for each segment of the sets obtained using the different segmentation algorithms. Finally, the classification process is performes with each of the other classification algorithm tested.

**Fig. 2** Flow chart of the classification process
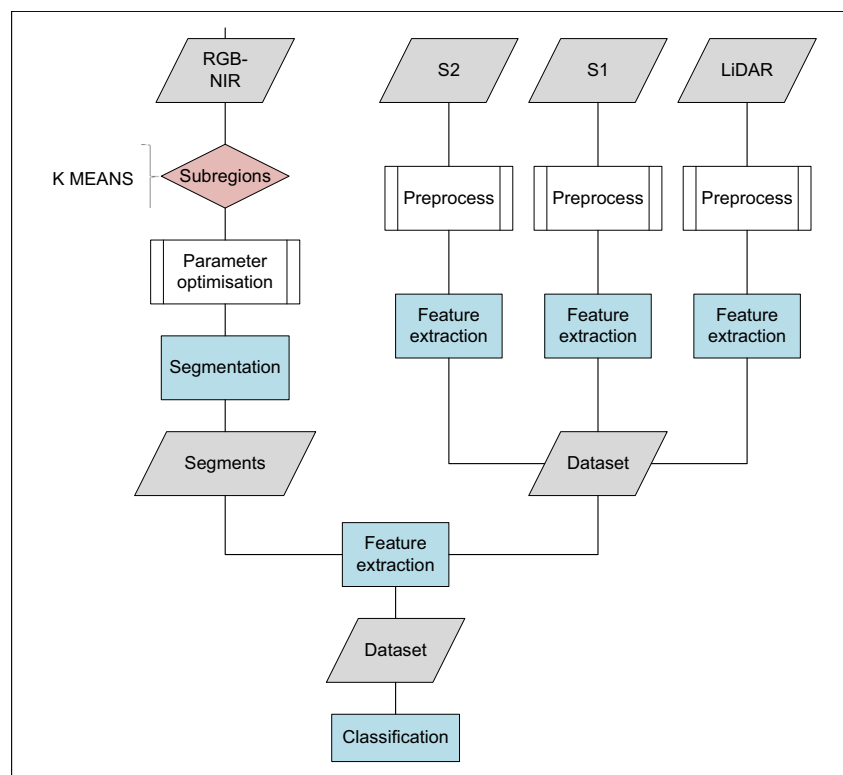
## Segmentation algorithms

Three algorithms based on similarity detection were used for segmentation: (i) Region Growing (RG), implemented in the Geographic Resources Analysis Support System (GRASS) GIS (Mitasova and Neteler 2008); (ii) Mean-Shift (MS), implemented in the ORFEO tools (McInerney and Kempeneers 2015); and (iii) Multiresolution Segmentation (MR) (Baatz and Schape 2000), implemented in TerraLib (Câmara et al. 2008).

### Region-growing

Region Growing (RG) is a robust but simple and popular method of segmentation (Adams and Bischof 1994). It starts by using each pixel within the selected region as an initial segment that joins adjacent pixels as long as they do not exceed a certain disimilarity threshold in the variable space set by the user. Once the process is complete, the algorithm forces the joining of segments or regions below a specified minimum size. Therefore, the parameters necessary for the operation of this algorithm are the similarity threshold (*threshold*) and the minimum segment size (*minsize*). It is available in GRASS GIS (Mitasova and Neteler 2008) through the *i.segment* module.

As with any algorithm, the problem is to determine the optimal parameters a priori. This problem was solved with a GRASS module created for this research which combines different values of both parameters and checks the results of 1000 segmentations. Their metrics were interpolated in the parameter space to find the optimal values.

### Mean-shift

This algorithm was not originally intended for image segmentation. It is a non-parametric method for estimating gradients in the density function (Fukunaga and Hostetler 1975). It was Comaniciu and Meer (2002) who applied it from a joint spatial and spectral point of view, grouping contiguous pixels whose modal estimation converges in the joint domain of both. This algorithm requires two parameters, a spectral range ($h_s$) and a spatial range ($h_r$) in which the pixels of an image converge in the local mode of the probability density to which they belong. The segmentation process ends when neighbouring pixels in the same mode are grouped together.

The optimisation process was similar to that used for the region-growing segmentation, where 1000 segmentations were performed and their metrics were interpolated in the parameter space.

### Multiresolution segmentation

The Multiresolution Segmentation (MS) algorithm (Baatz and Schape 2000) is a hierarchical *region growing* type of technique where objects are grouped, starting from individual pixels, up to a threshold that marks the maximum variance between objects. It requires four input parameters: the scale parameter refers to the variance threshold (*threshold*), which in turn is weighted by the shape parameters (*colour* and *compactness*); finally, it also requires a minimum size parameter (*minsize*).

The optimisation was carried out using the same strategy as the others, increasing the number of trials to 4000 due to the greater number of parameters to be determined. It should be noted that during the execution of the segmentation algorithms, the MR algorithm experienced problems in execution, which were solved by dividing each region into four sub-regions, keeping the parameters of the original region for all four, and then rejoining them to form the final map.

However, when analysing the results of the inverse distance interpolation in 4 dimensions, the compromise solution was not as clear as for the other algorithms. Figure 3 shows the values of the final metric against the logarithm of the number of polygons generated. As can be seen, the lower the number of polygons generated by the segmentation, the lower the final metric. On the other hand, there are many segmentations with very similar high metrics, but with a large variation in the number of polygons obtained as a result. Both the optimisation and the segmentation using this algorithm were carried out using the SegOptim library of R (Gonçalves et al. 2019), which is linked to the TerraLib software (Câmara et al. 2008).
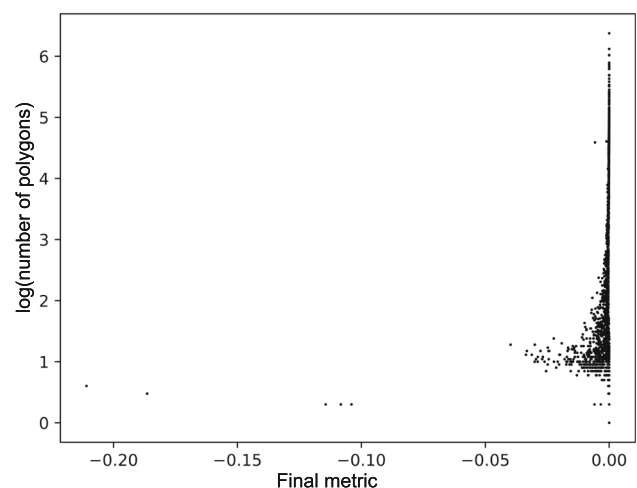


**Fig. 3** Results of the optimization of the segmentation parameters with TERRALib for region 1

**Fig. 4** Clusters obtained by k-means with final delimitation and region identifier
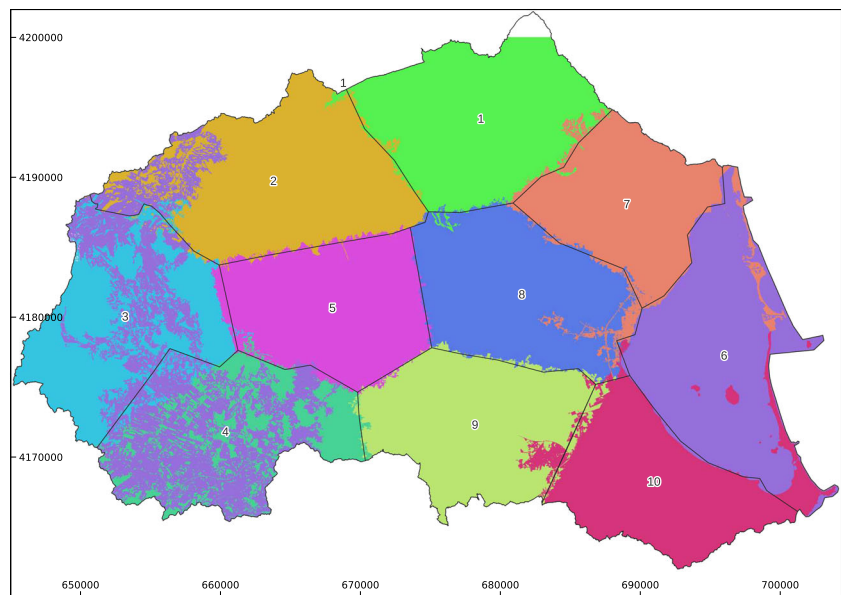


## Region pre-segmentation

The most common problem in image segmentation is related to the choice of parameters for object recognition. Since the study area may contain different regions with different types and scales of objects, the area was first subdivided. For this purpose, land cover polygons resulting from a previous pixel-based multisensor classification (Valdivieso-Ros et al. 2023) were regrouped according to the location of each centroid and their size using the k-means method, after estimating the optimal number of clusters to be 10.

To complete the process of subdividing the basin into smaller regions, the boundaries of the spatial clusters obtained by k-means were redefined to homogenise the polygons defined by the algorithm, resulting in the delineation of the 10 regions shown in Fig. 4.

The parameter optimisation and segmentation process was performed for each region with each segmentation algorithm. The Moran index and variance of 1000 different parameter combinations were calculated, and an objective function integrating both metrics was generated. This was followed by an interpolation procedure in the parameter space, from which a compromise solution of parameter values was taken for each region. This was followed by segmentation of the regions from the RGBI band combination. Finally, the segmentation maps were merged to calculate the geometric information and the statistics of the predictor variables of each segment. These data form the predictor sets for classification.

## Datasets generated from each segmentation

Six datasets were prepared. For each of the three segmentation algorithms, the variables proposed in Cánovas-García

and Alonso-Sarría (2015) were calculated. These include a set of geometric and morphological variables of the objects: Length, area, width, perimeter, edge index, compacity, fit to an ellipse, fit to a rectangle and circularity. These variables were combined with statistics derived from the most relevant variables in a previous pixel-based classification (Table 1). Means and standard deviations were computed in one case, and medians and ranges in the other. This resulted in 6 datasets, each containing 262 features. The description of the features can be found in Valdivieso-Ros et al. (2023).

**Table 1** Summary of datasets and features selected in Valdivieso-Ros et al. (2023)

| Dataset | Variables |
| --- | --- |
| S1 | VV, VH |
| S1 indices | DPSVI |
| S2 | B01, B02, B03, B04, B05, B06, B07, B08, B08A, B11, B12 |
| S2 indices | NDVI, SAVI, NDBI, MNDWI, TCB |
| S2 texture | PC1, NDVI, Entropy, Contrast, Second angular moment |
| LiDAR | ppA, ppM, ppB, ppH, ppE, mZG, mZB, mZM, mZH, mZA, mZE, sZG, sZB, sZM, |
|  | sZA, sZE, sZH, Hv, He, Nk, Nke, Nvv, wCv, wCd, wDv, wDd |

The LiDAR date is 2018/08/01; Sentinel2 dates are 2018/11/07, 2019/02/25, 2019/04/11, 2019/06/10; Sentinel1 dates are 2018/11/08, 2019/02/24, 2019/04/13 y 2019/05/19

The Sentinel-1 and Sentinel-2 dates were chosen to be as close as possible between sensors and spread across the different seasons of the year to best match the most relevant crop ephemerides in the study area (Valdivieso-Ros et al. 2023)

## Training segments and classification scheme

A training dataset was obtained from the segments produced by each of the three segmentation algorithms, avoiding overlap with training polygons used in a previous pixel-based classification (Valdivieso-Ros et al. 2023) in order to use these polygons as a test set for the final classifications of the polygons produced by each algorithm. The training segments were manually selected based on the most recent ortophotography of the National Plan of Aereal Ortopotography (PNOA, by its initials in Spanish) of 2019, with a stratified sampling design to ensure a reasonable presence of segments of all classes containing only one, although this condition was not always easy to meet. Table 2 shows the classification scheme, the number of segments for each class in the three training datasets, and the number of pixels for the final test validation.

Figure 5 shows the location and geometry of the training segments selected from those obtained by each segmentation algorithm.

## Classification algorithms

Classifications were performed with two algorithm, Random Forest (RF) and Multilayer Perceptron (MLP), validating the models by LOO-CV cross-validation. In addition, we checked the suitability of other supporting validation of the accuracy results by performing a confusion matrix between the final predictions of each GEOBIA model and an indepedent set of training pixels.

### Random forest

Random Forest (RF) (Breiman 2001) is an ensemble learning method for non-parametric classification, based on decision trees. It reduces the correlation between trees through two modifications that better fit to the concept of ensemble (James

et al. 2013; Gomariz-Castillo et al. 2017): training each tree with a subset of the training data and selecting a random subset of predictors for each node. The class of each pixel in an image is assigned based on the majority vote of the trees, with the option to obtain uncertainty metrics.

RF has several advantages (Liaw and Wiener 2002; Rodriguez-Galiano et al. 2012; Belgiu and Drăgu 2016; Cánovas-García et al. 2017; Maxwell et al. 2018), including robustness in cases of high dimensionality, correlated variables, small training datasets, or anomalous data. It can also provide an importance ranking of variables and can be used as an explanatory model and an internal goodness of fit metric using the out-of-bag technique (OOB-CV). In addition, it requires less computation than other machine learning methods and generates an internal metric of validation using the out-of-bag technique (OOB-CV), which is an unbiased estimation of the generalisation error as the training data are randomly sampled.

For this study, the number of trees was set to 2000, and a two-stage variable selection using the Variance Inflation Factor (VIF) and the importance ranking of the variables of the model (Valdivieso-Ros et al. 2023) wass applied to deal with the multicollinearity of the variables and the high dimensionality. The RF model was carried out using the package *randomForest* (Liaw et al. 2014), available for the R software (Team 2020).

### Multilayer perceptron

A Multilayer Perceptron (MLP) is an artificial neural network that uses multiple layers of interconnected neurons and non-linear activation functions to learn to perform specific tasks through a supervised training process. It consists of at least three layers of neurons: an input layer, one (or more) hidden layers and the output layer. The layers of neurons are connected by a mathematical function (activation function).

**Table 2** Classification scheme, including the number of training and validation segments of every dataset from every algorithm: Region Growing (RG), Mean-Shift (MS) and Multiresolution (MR), as well as the number of pixels for every class in pixel test validation dataset

| Id | Class | Description | Segments | | | Test pixels |
|----|-------|-------------|------|------|------|------------|
| | | | RG | MS | MR | |
| 1 | Forest | Mediterranean forest | 42 | 27 | 48 | 1000 |
| 2 | Shrubland | Low vegetation | 40 | 31 | 53 | 1200 |
| 3 | Dense woodland | Fruit and citrus crops | 42 | 40 | 57 | 1800 |
| 4 | Irrigated arable crops | Horticultural crops | 40 | 41 | 39 | 1000 |
| 5 | Impermeable | Artificial Surfaces | 44 | 38 | 36 | 1639 |
| 6 | Water | Water bodies, including Artificial reservoirs | 41 | 24 | 41 | 1158 |
| 7 | Bare soil | Bedrock without cover or sparsely vegetated | 41 | 26 | 20 | 1055 |
| 8 | greenhouses | Irrigated areas covered with plastic | 30 | 29 | 29 | 2600 |
| 9 | Netting | Irrigated crops covered with netting | 19 | 17 | 13 | 1400 |
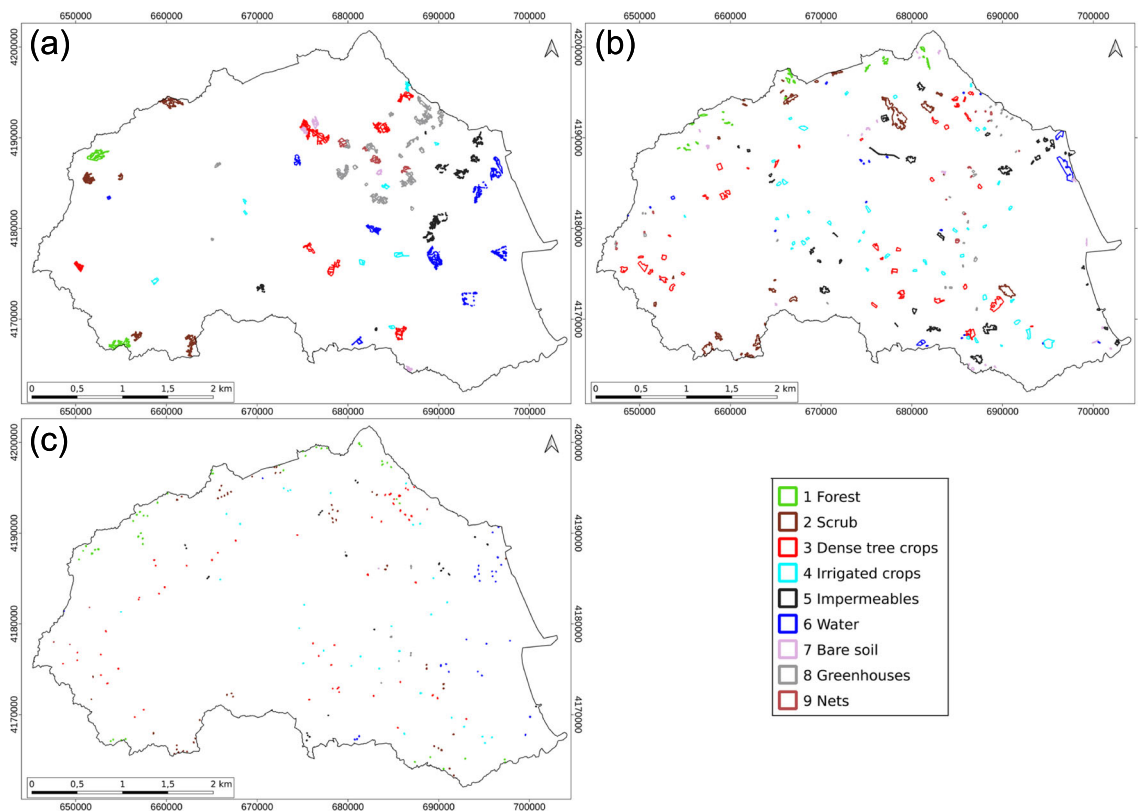
**Fig. 5** Maps of training and validation segments distribution of every segmentation algorithm dataset. (a) Region-growing (rg), (b) Mean-Shift (MS) and (c) Multiresoution (MR)

The input data is transmitted as a linear combination to which the activation function is applied and its output is redirected to the next layer, following the same computational pattern. The final layer, or output layer, produces the output of the network, which can be a number, a vector or a matrix, depending on the type of problem being solved. The output is computed in a similar way to the hidden layers, but with a different activation function.

The goal of learning in an MLP is to adjust the weights of the connections between the nodes so that the network can perform a specific task, such as classification, through supervised training. This training consists of providing the network with an input along with the expected output so that the network adjusts the weights to minimise the error between the predicted output and the expected output.

The neural network used in this study consists of an input layer with the predictors, three dense hidden layers of 300, 100 and 50 neurons, and an output layer of 9 neurons, one for each possible output class. A ReLU (*Rectified Linear Unit*) activation function (Hahnloser et al. 2000) was applied to each hidden layer, and a normalised exponential function (*softmax*) to the output, as this allows each output to be

interpreted as a probability for each class, guaranteeing that the sum of all probabilities is equal to one. An Adaptative Moment Estimation (ADAM) was chosen as the gradient optimisation algorithm. This algorithm was proposed by Kingma and Ba (2015) and is commonly used to train neural networks. In addition, it has the advantage of also having an internal adaptive method for setting the best learning rate. The *dropout* or *dropout rate* is a technique to prevent overfitting of neural network models by temporarily deactivating some randomly selected neurons during each training iteration. Testing different values of the drop rate allows to avoid the problems of overfitting due to an overparameterised network (several layers with many neurons) without having to change the actual architecture of the network.

The number of neural network iterations (*epochs*) is 100, although the early stopping technique (*callback early stopping*) is used by setting a patience threshold of 20. If the accuracy does not improve in this number of epochs, no further epochs are run and the best model is selected from those already obtained. In this study, the tensorflow (Abadi et al. 2015) python package was used to build and train the neural networks.

## Validation

The validation of the models was done in two ways, first with a cross-validation procedure leaving out one polygon at a time (LOPO-CV) and later by calculating a confusion matrix of the segment prediction datasets compared to the pixel dataset used as a reference for the selection of the training areas of the aforementioned pixel-based classification (Valdivieso-Ros et al. 2023). The latter method could provide insight into the performance of the classification but also into the segmentation process.

## Results

### Segmentation

Table 3 shows a compromise solution from the optimisation processes. Apart from the number of parameters, two for RG and MS but four for MR, there are significant differences in the values from one algorithm to another. However, the parameters of each algorithm measure different aspects of the process to determine the optimal segmentation of the image, making them difficult to compare. It is also worth noting that the optimisation processes were time and resource consuming, being higher and more complex for MR due to the number of parameters to optimise.

Table 3 shows, per region, the total number of segments obtained with each segmentation algorithm. The number of segments per region refers to the full extent of each region. The final number of segments in the study area, after applying the masks to each region and merging them, is shown in the Total column.

MS is the algorithm that produces the smallest number of segments and with the largest size, as can be seen as an example in Fig. 6. In some regions the three algorithms agree

in calculating smaller segments and therefore a larger number of segments compared to the other regions, as in region 3. A similar problem occurred in region 6, whose area is mainly occupied by water, as it includes the lagoon and its coastline.

From a qualitative point of view, Fig. 6 shows an extract of the segment maps obtained, which are clearly different in size, with those calculated by the MS algorithm being the largest. The figure shows how these segments tend to mix objects belonging to different classes, such as irrigated grass plots or buildings next to plots with and without vegetation. There is clearly an under-segmentation, whereas in the examples corresponding to the other two algorithms, there is a clear over-segmentation; this is clear in both, but in the RG segments more concentrated in the urban area. The segments of the RG and MS algorithms also show a more varied geometric segmentation of different sizes. The appearance of linear objects such as roads is noticeable, although sometimes mixed with neighbouring objects such as ponds. In contrast, the MR algorithm does not seem to have resolved these geometric differences well, producing much smaller objects that do not seem to respond well to an over-segmentation of larger objects, as they would include part of the neighbouring objects when joined.

### Classifications with RF

Following the two-step variable selection, the dimensionality of the datasets is significantly reduced from the original 262 variables (Table 4). The first VIF-based step reduces the number of predictors by about 50 % in all cases. The second step, based on RF importance, generally leads to a greater reduction, but with greater variability.

The accuracy metrics per algorithm, dataset and validation type are shown in Table 5 with their 95 % confidence intervals. The segmentation produced by the MR algorithm has the highest accuracy values, but the differences are only

**Table 3** Optimal parameters estimated and number of segments obtained for each region with each algorithm

| Region | RG | | | MS | | | MR | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Thres. | Minsize | Nseg. | $h_s$ | $h_r$ | Nseg | Compact. | Color | Minsize | Thres. | Nseg |
| 1 | 0.05 | 25 | 6729 | 0.05 | 6 | 119643 | 0.032 | 0.022 | 27 | 0.003 | 47260 |
| 2 | 0.05 | 6 | 14691 | 0.035 | 4 | 120575 | 0.075 | 0.025 | 38 | 0.004 | 47090 |
| 3 | 0.01 | 2 | 225702 | 0.045 | 6 | 140655 | 0.031 | 0.031 | 15 | 0.002 | 102417 |
| 4 | 0.05 | 2 | 23111 | 0.025 | 6 | 104848 | 0.015 | 0.092 | 40 | 0.025 | 33167 |
| 5 | 0.01 | 6 | 44721 | 0.035 | 5 | 81364 | 0.092 | 0.082 | 16 | 0.012 | 24550 |
| 6 | 0.02 | 4 | 94207 | 0.05 | 9 | 82212 | 0.058 | 0.018 | 28 | 0.006 | 59781 |
| 7 | 0.05 | 2 | 27585 | 0.045 | 6 | 95616 | 0.934 | 0.623 | 12 | 0.262 | 23914 |
| 8 | 0.05 | 10 | 8933 | 0.045 | 8 | 101361 | 0.045 | 0.020 | 30 | 0.005 | 23725 |
| 9 | 0.05 | 10 | 10504 | 0.4 | 6 | 95835 | 0.074 | 0.030 | 25 | 0.007 | 45697 |
| 10 | 0.05 | 2 | 45051 | 0.4 | 4 | 53438 | 0.090 | 0.061 | 46 | 0.005 | 28308 |

**Fig. 6** Segments obtained with each algorithm in a peri-urban agricultural area



Multiresolution segmentation

Mean shift

Region growing
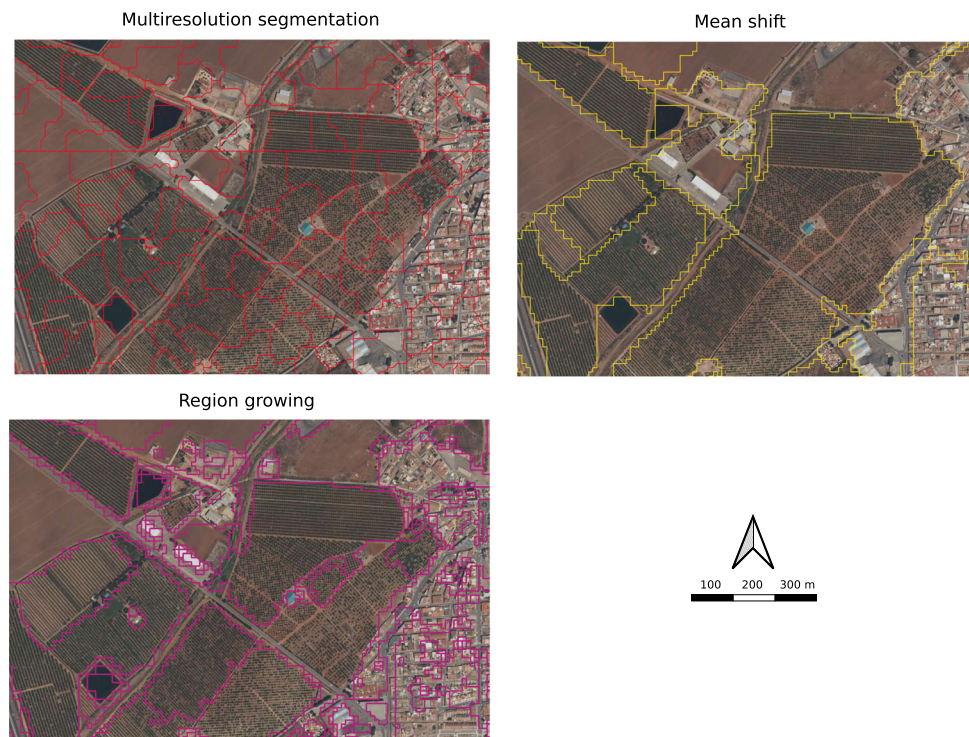
significant in the pixel test not in the LOPO-CV test. The differences between using mean and standard or using median and range are not significant. Figure 7 clearly shows that there are no significant differences between the different accuracy metrics, according to their confidence intervals, except between the datasets of mean+stdev of MS and MR, which is the one with the lowest confidence intervals.

By class, there is more variation in the accuracy of the different models (Table 8 in Appendix A). The segment sets obtained with RG give better results for all classes, although it is not the best overall model. On the contrary, the segment sets of the MR algorithm present serious problems for the classification of class 9 (netting), being the models that present the best global accuracies.

The obtained omission and commission errors between classes (Table 9 in Appendix A) are very high for greenhouses and nets in almost all datasets. Figure 8 shows the

**Table 4** Number of variables remaining after each stage of the selection process

|  | Data set | VIF selection | Final predictors |
|---|---|---|---|
| RG | mean+stdev | 127 | 27 |
|  | median+range | 141 | 20 |
| MS | mean+stdev | 127 | 16 |
|  | median+range | 185 | 162 |
| MR | mean+stdev | 137 | 61 |
|  | median+range | 131 | 20 |

confusion errors of the best-accuracy dataset for each segmentation algorithm.

The natural vegetation classes are best classified with the MS median+range dataset, while the other classes are more evenly distributed between the RG and MR datasets. The algorithm that performs best with the most problematic vegetation class of irrigated crops is RG, in principle with mean+stdev, but with median+range following closely and above the others. In this class, irrigated crops, the diversity of crop types and the fragmentation of the fields are a problem to be solved for the classification of land use in the area. The highest errors of commission and omission in all the classifications are concentrated in class 9, well above 0.2, which also exceeds class 8 (greenhouses) in all classifications and class 7 (bare soil) in some.

The usual confusion in identifying and predicting the classes of nets for greenhouses and dense tree crops can be explained by the fact that, on the one hand, the materials from which these nets are made are similar to those used for some types of greenhouse cover in this area and, on the other hand, they are mainly used as protective cover for tree crops.

Regarding the final maps, the differences are mostly found between MR and the other algorithms (Fig. 9). The MR accuracy is the highest, and it can be concluded that the classification tends to predict the predominant class in the segment, even when the segments do not clearly correspond to an object (plots, ponds or greenhouses, among others), nor when they are combined with adjacent ones.

**Table 5** Accuracy of RF classifications by dataset and segmentation method with 95% confidence intervals. The dataset with the highest accuracy per algorithm is highlighted in bold

| Validation type | Data set | Accuracy | CI 95% | Kappa Index | CI 95% |
|---|---|---|---|---|---|
| LOPO-CV | RG mean+stdev | 0.867 | 0.826, 0.902 | 0.850 | 0.809, 0.890 |
| | **RG median+range** | 0.870 | 0.830, 0.904 | 0.853 | 0.813, 0.893 |
| | MS mean+stdev | 0.802 | 0.750, 0.848 | 0.775 | 0.722, 0.828 |
| | **MS median+range** | 0.861 | 0.814, 0.899 | 0.842 | 0.795, 0.888 |
| | **MR mean+stdev** | 0.905 | 0.868, 0.934 | 0.891 | 0.855, 0.926 |
| | MR median+range | 0.893 | 0.855, 0.924 | 0.877 | 0.839, 0.915 |
| Pixels Test | **RG mean+stdev** | 0.926 | 0.924, 0.928 | 0.907 | 0.904, 0.909 |
| | RG median+range | 0.916 | 0.915, 0.918 | 0.895 | 0.893, 0.897 |
| | MS mean+stdev | 0.886 | 0.884, 0.888 | 0.857 | 0.854, 0.859 |
| | **MS median+range** | 0.905 | 0.903, 0.907 | 0.881 | 0.878, 0.883 |
| | **MR mean+stdev** | 0.930 | 0.928, 0.932 | 0.912 | 0.910, 0.914 |
| | MR median+range | 0.930 | 0.927, 0.931 | 0.910 | 0.908, 0.912 |

Conversely, larger objects can be identified despite the under-segmentation produced by the MS algorithm, e.g. linking ponds and buildings with agricultural plots in cultivated areas. A similar situation can be observed with the RF classification of the RG dataset. An example is shown in Fig. 10.

Figure 11 shows prediction maps of the whole basin obtained with the best dataset of each segmentation algorithm. Differences in the global surfaces of the classes can be observed; RG and MR are qualitatively more accurate, without taking into account the problem of classifying greenhouses and netting. Classification with datasets from both algorithms shows a high tendency to confuse these classes, not only among themselves, but also with both cultivated

classes (dense tree crops and irrigated grass crops) and impermeable which is translated in the final map into observable differences in the predicted total areas of crop classes: while in the RG dataset prediction with RF dense tree crops and scrub seem to be the classes with more area, in MR irrigated crops and greenhouses area is higher. There may also be differences in the total area of impermeable surfaces. It is mainly confused with bare soil in both predictions. However, omission and commission errors in the RG classification of impermeable segments are distributed over a larger number of classes (greenhouses, netting, bare soil, water and scrub). The MR problems in detecting and predicting bare soil shown in Fig. 8, are reflected in less area of this class in the MR map than in the maps of RG and MS.

**Fig. 7** Accuracy metrics and confidence intervals of the final rankings with test pixels
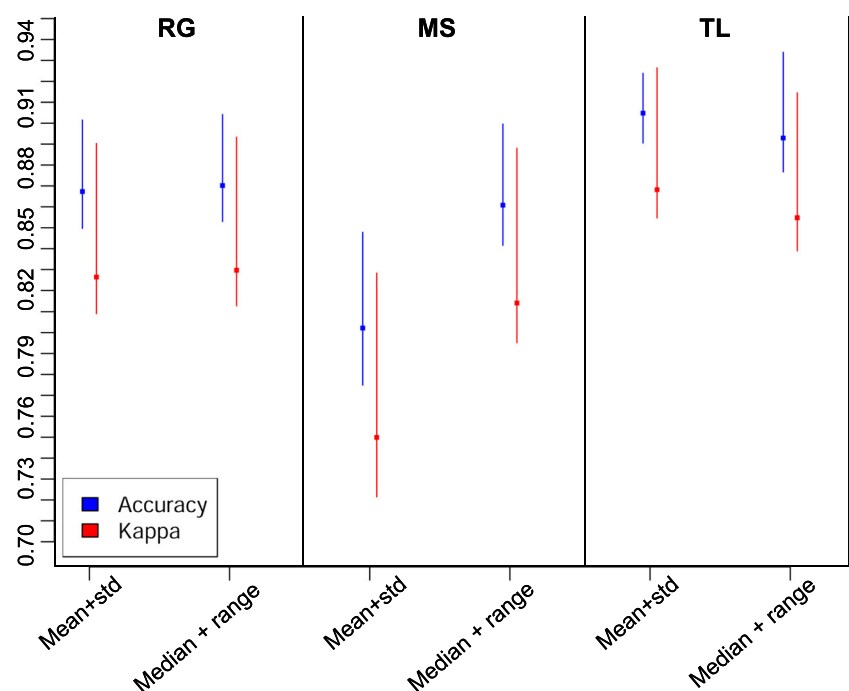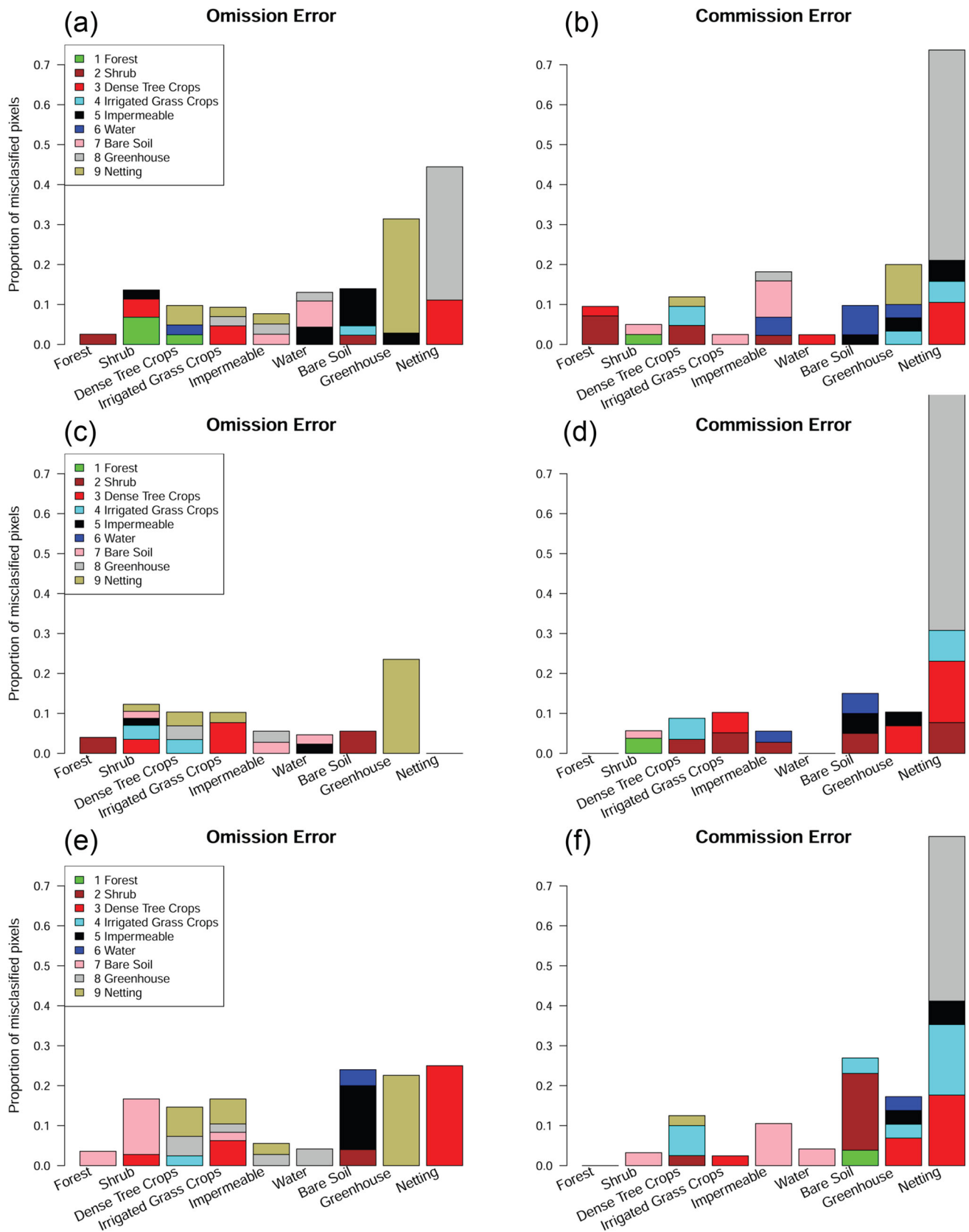
**Fig. 8** Cross-class confusion plots of the best-accuracy datasets for each segmentation algorithm: (a-b) RG mean+stdev, (c-d) MR mean+stdev and (e-f) MS median+range
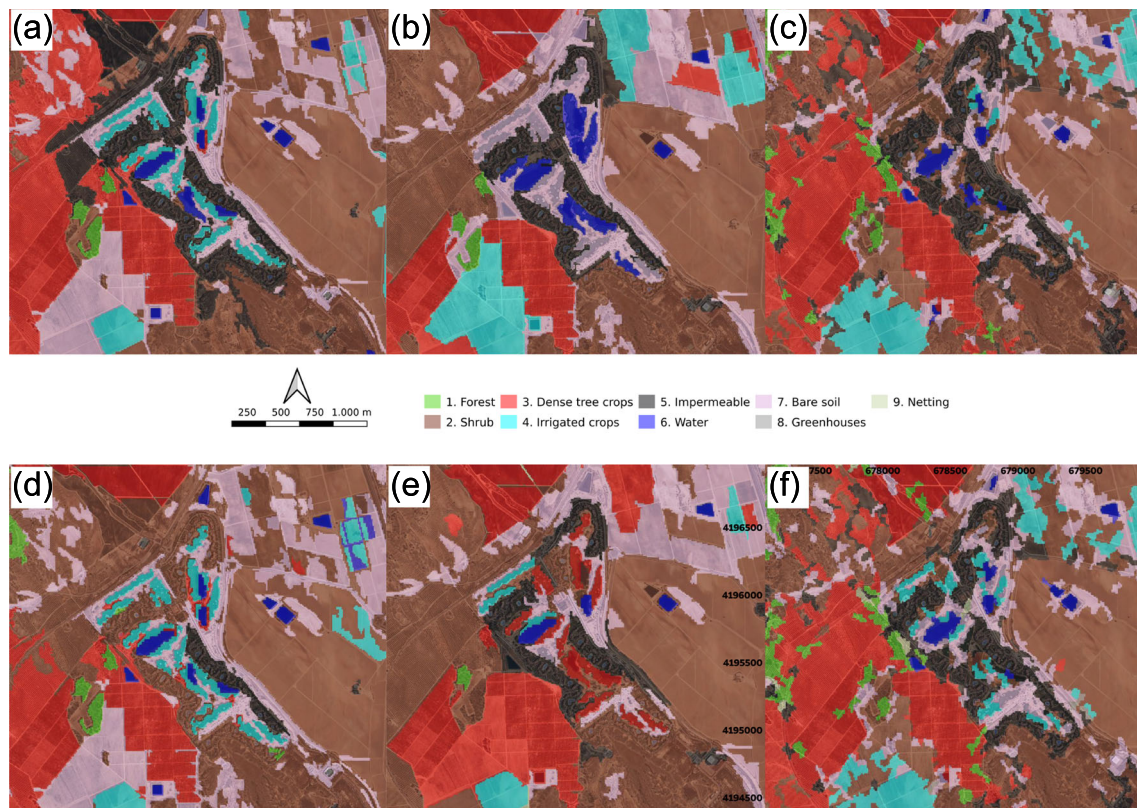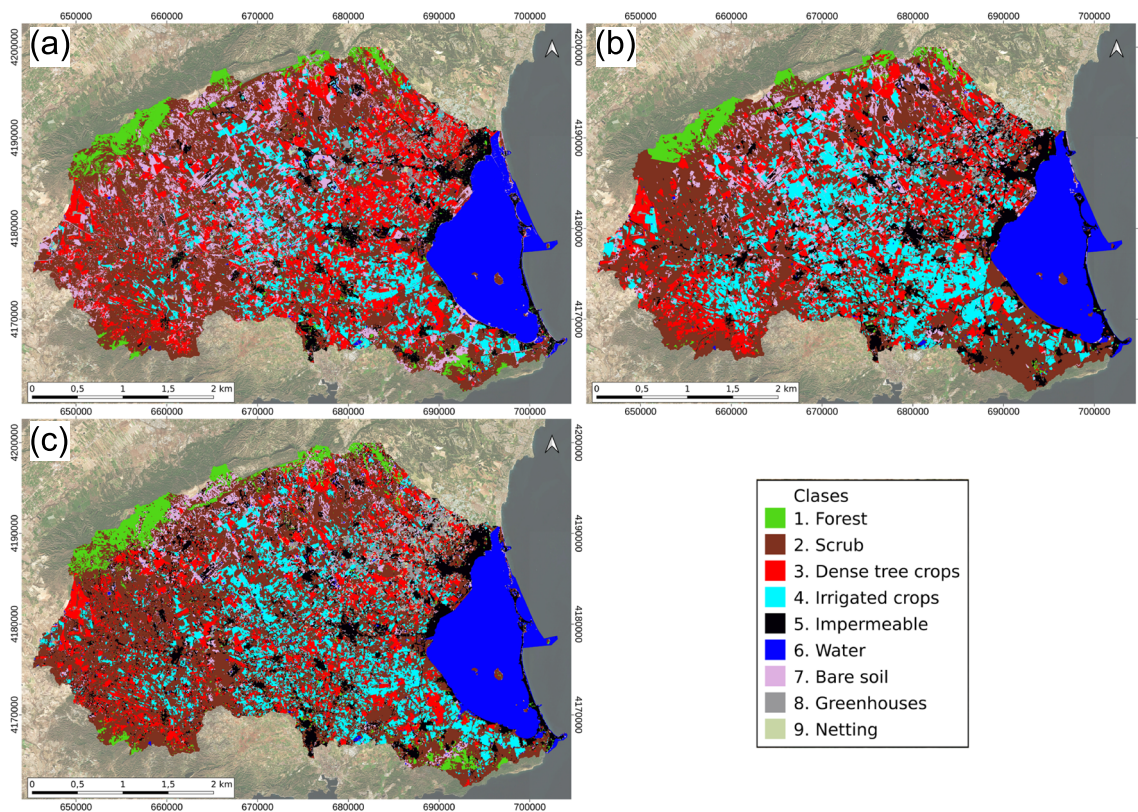
**Fig. 9** Comparison of classification results in a residential area surrounded by crops parcels with RF. (a,d) RG mean+stdev and median+range, (b,e) MS mean+stdev and median+range, (c,f) MR mean+stdev and median+range

**Fig. 10** Comparison of classification results in agricultural area with RF. (a) RG, (b) MR and (c) MS
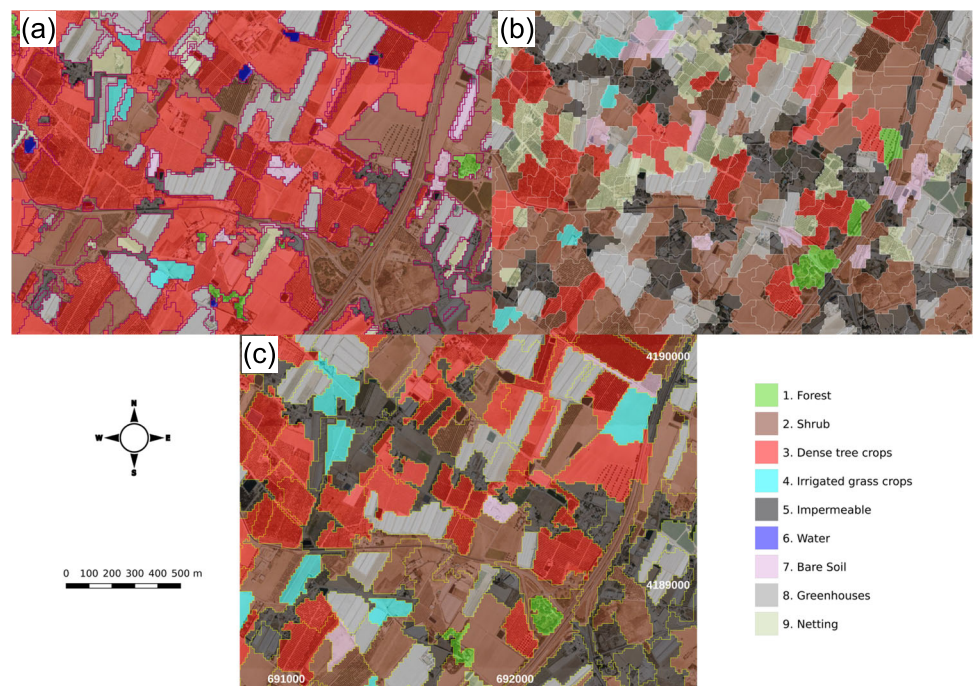
**Fig. 11** Final maps obtained with best models by algorithm with RF. (a) RG, (b) MS and (c) MR

## Classification with MLP

Tables 6 and 7 show the accuracy metrics obtained by the MLP model per segmentation algorithm, dataset and validation type. In this case, several values of the drop rate were tested.

Table 6 shows that there are no clear differences in accuracy when the drop rate is changed, except for the maximum value tested (0.8). This means that although the model was probably over-parameterised, this over-parameterisation did not result in a clearly over-trained model. Table 7 shows that the MR segmentation model using the mean and range of the spectral and LiDAR features achieves significantly better results than the other methods. The RG model appears in second place with no significant differences between using mean and standard deviation or median and range. In general, there are no significant differences between the types of statistics used, but the differences between the segmentation methods are significant.

Analysing the performance by classes, there are large differences between models (Table 10 in Appendix A). There is no single dataset that shows better classification results for all classes. In fact, they all show classification problems, especially for the netting class.

The highest errors of omission and commission occur in greenhouses and netting (Table 11 in Appendix A), which are very susceptible to confusion within each other and with other related classes such as dense tree crops, irrigated crops and impermeable, as can be seen in the graphs in Fig. 12. Both natural vegetation classes are well classified with the MS dataset, while MR only correctly classifies the forest class. The problem of confusion between the two vegetation types is not solved by any of the segmentation algorithms, although that of RG seems to solve it more efficiently. However, the remaining omission and commission error rates obtained with this dataset are similar or even higher than those obtained with the MS or MR datasets. Interestingly, the variety of confusions between classes obtained with the MR algotithm seems to be reduced to a group of classes dominated by soil (with no, low or sparse vegetation as scrub, impermeable and bare soil), and another group dominated by cultivated vegetation (both classes of crops, greenhouses and netting).

There are large differences between the maps produced by MLP and those obtained with RF. Qualitatively, the worst classification results coincide with the quantitave results: The map produced with the MS dataset had the lowest qualitative accuracy (Fig. 13). In the area predicted by the RG dataset, the prevalence of dense tree crops is noticeable, as is the

**Table 6** Accuracy of MLP classifications by dataset and segmentation method

| Data set | Drop Rate | Accuracy | Test Accuracy |
|---|---|---|---|
| RG mean+stdev | 0.0 | 0.8642 | 0.8348 |
| | 0.2 | 0.8658 | 0.8466 |
| | 0.4 | 0.8928 | 0.8720 |
| | **0.6** | 0.8883 | **0.875** |
| | 0.8 | 0.5228 | 0.4702 |
| RG median+range | 0.0 | 0.8563 | 0.8289 |
| | **0.2** | 0.8625 | **0.8407** |
| | 0.4 | 0.8613 | 0.8407 |
| | 0.6 | 0.8513 | 0.8259 |
| | 0.8 | 0.3219 | 0.2507 |
| MS mean+stdev | 0.0 | 0.8698 | 0.8242 |
| | **0.2** | 0.8652 | **0.8461** |
| | 0.4 | 0.8592 | 0.8278 |
| | 0.6 | 0.8365 | 0.7985 |
| | 0.8 | 0.3410 | 0.2821 |
| MS median+range | 0.0 | 0.8607 | 0.8425 |
| | 0.2 | 0.8645 | 0.8242 |
| | 0.4 | 0.8617 | 0.8095 |
| | **0.6** | 0.8427 | **0.8425** |
| | 0.8 | 0.3218 | 0.2344 |
| MR mean+stdev | **0.0** | 0.9002 | **0.8958** |
| | 0.2 | 0.8986 | 0.8780 |
| | 0.4 | 0.8928 | 0.8720 |
| | 0.6 | 0.8883 | 0.875 |
| | 0.8 | 0.5228 | 0.4702 |
| MR median+range | **0.0** | 0.8906 | **0.875** |
| | 0.2 | 0.8959 | 0.8661 |
| | 0.4 | 0.8963 | 0.8690 |
| | 0.6 | 0.8842 | 0.8720 |
| | 0.8 | 0.5034 | 0.4494 |

The dataset with the highest test data accuracy per algorithm is highlighted in bold, along with the dropout rate at which it is achieved

case with the RF dataset, while the MR dataset predicts irrigated crops. In this dataset, in the area where greenhouses and nets actually dominate, the former were mostly labelled.

However, in general, the separability is really low, not only between spectrally similar classes, the diversity of crop types and the fragmentation of the landscape, a real problem of land use and land cover classification in the area, are not handled as well as other ML algorithms with any other dataset.

It is noteworthy the case of the south-eastern part of the catchment, where some areas are clearly under-segmented and misclassified as water or bare soil with the MS dataset (Fig. 14).

## Discussion

The qualitative analysis of the 3 segmentations obtained shows that the 10 meter resolution of Sentinel-2 is probably not sufficient for segmentation in this particular study area because of the fragmentation of the landscape. A higher spatial resolution is required. The intra-class variability in some of the classes of the classification scheme, which is expected to be very high, poses a problem similar to that pointed out by Blaschke et al. (2008, 2014), among others, when referring to the problem of pixel-based classifications. They also concluded that, in an urban context, there is no optimal scale for identifying different objects and that each algorithm deals with the scale problem in a different way, a statement that can be observed in the results obtained in this study. One of the peculiarities of this study area is the presence of small urban-like structures scattered across a predominantly agricultural landscape. The RG and MR segmentation algorithms have solved the problem more efficiently in qualitative terms, achieving better quantitative results in the global classification metrics, while MS undersegmented the image, with the worst results in both qualitatively and quantitatively terms.

Hossain and Chen (2019) identifies good performance with low computational cost as an advantage of RG-type algorithms. Certainly, in this study, the overall accuracy, although lower than that obtained with the MR algorithm, it is not quite lower in either of the two datasets extracted from its segments. Furthermore, it is implemented in an open software GIS and can be easily optimised. It is the most accurate of the three in terms of class classification accuracy using

**Table 7** Pixel test accuracy of MLP classifications by dataset and segmentation method with 95% confidence intervals

| Predictors | Accuracy | CI 95% | Kappa Index | CI 95% |
|---|---|---|---|---|
| RG mean+stdev | 0.924 | 0.922, 0.926 | 0.904 | 0.902, 0.906 |
| **RG median+range** | 0.925 | 0.923, 0.926 | 0.905 | 0.903, 0.907 |
| MS mean+stdev | 0.894 | 0.892, 0.896 | 0.866 | 0.863, 0.868 |
| **MS median+range** | 0.902 | 0.899, 0.904 | 0.876 | 0.873, 0.878 |
| **MR mean+stdev** | 0.945 | 0.943, 0.946 | 0.930 | 0.928, 0.932 |
| MR median+range | 0.913 | 0.911, 0.915 | 0.890 | 0.888, 0.892 |

The dataset with the best accuracy per algorithm is highlighted in bold

both RF and MLP, especially using RF, even for those classes with similar geometric and radiometric characteristics in the segments. This aspect was highlighted in the review by Hossain and Chen (2019), when analysing similar techniques to determine their applicability, as one that needs further investigation. In the same study, the segmentation of linear objects

such as roads or paths was identified as a key challenge due to their spectral similarity to, for example, some roofs or turbid water. In this study, both RG and MS algorithms were found to properly extract linear objects, albeit in a fragmented manner, although MS did associate them with water bodies such as irrigation ponds at their edges.
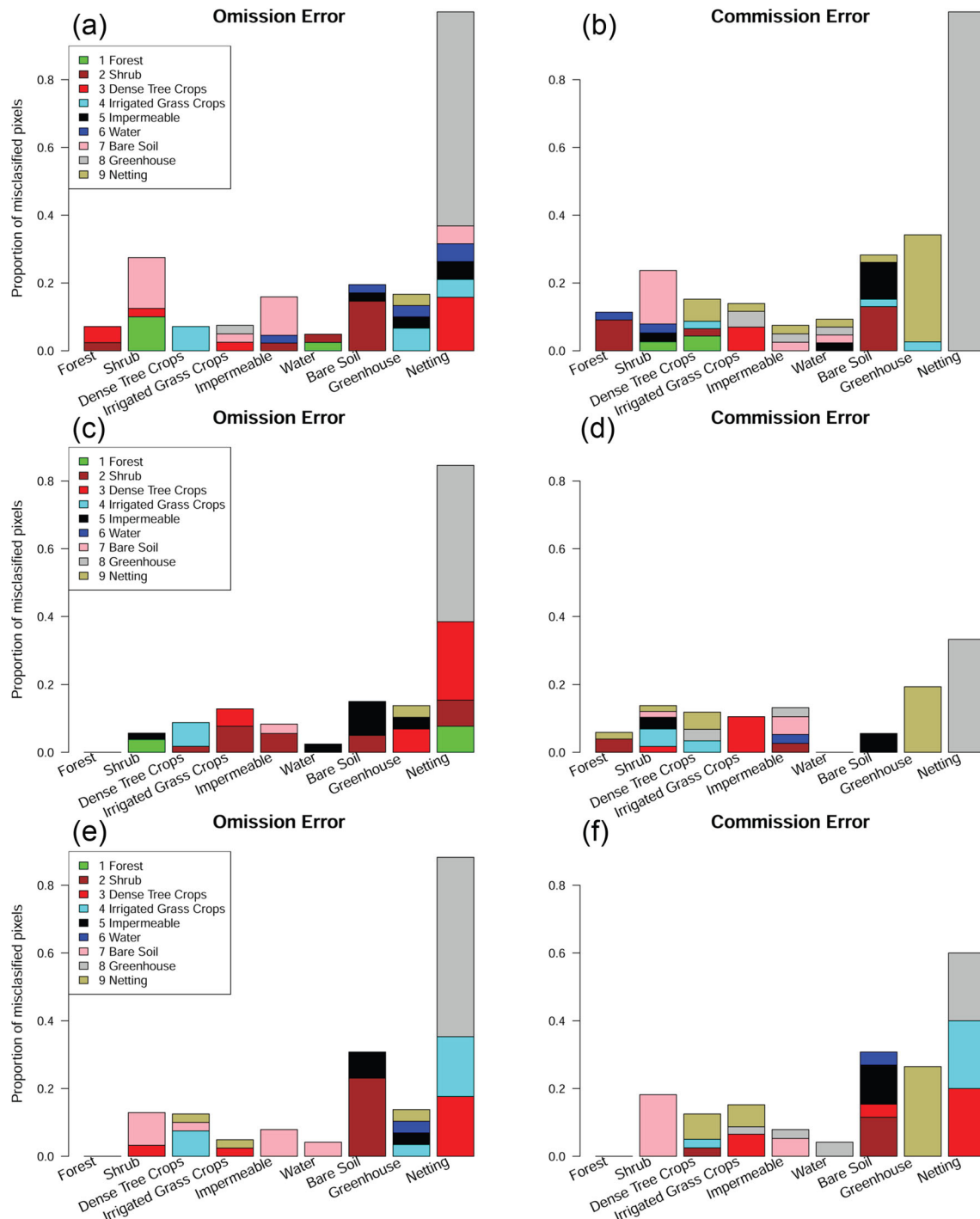


**Fig. 12** Cross-class confusion plots of the best-acccuracy datasets for each segmentation algorithm: (a,b) RG mean+stdev, (c,d) MR mean+stdev and (e,f) MS mean+stdev
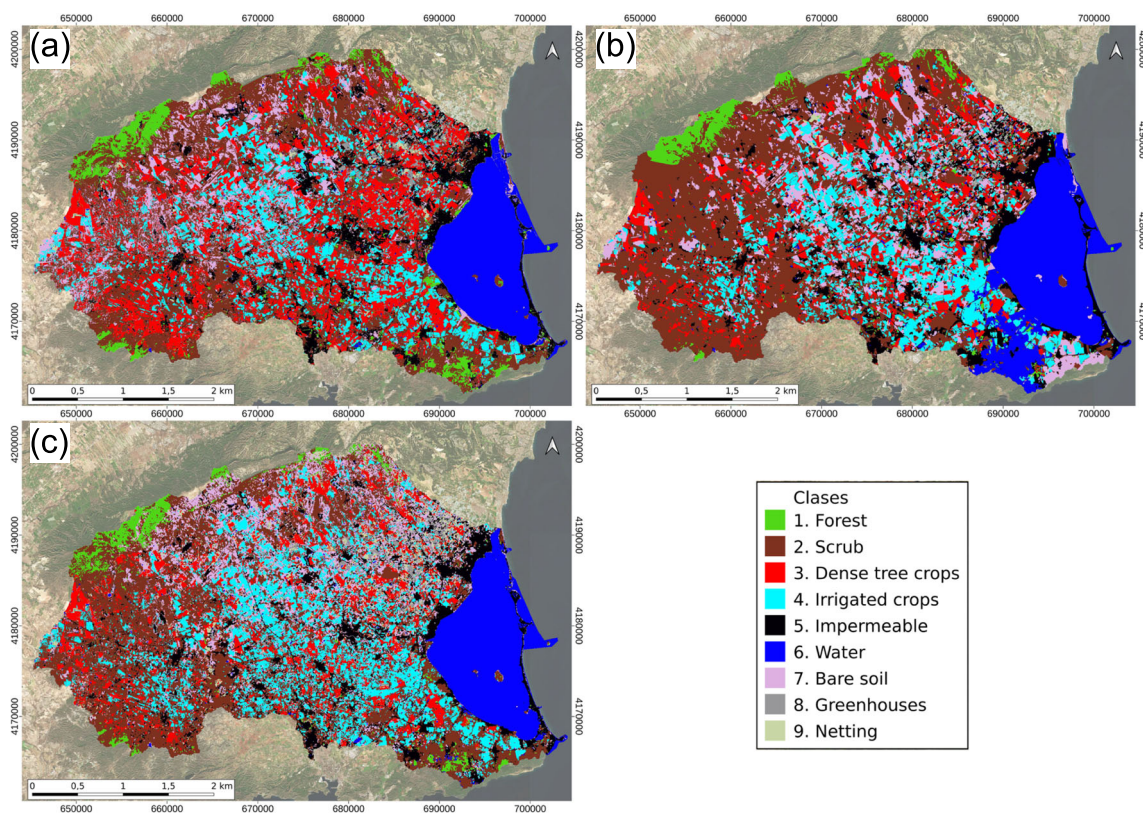
**Fig. 13** Final maps obtained with better MLP datasets of every segmentation algorithm: RG mean+stdev (a), MS mean+stdev (b) and MR mean+stdev (c)

**Fig. 14** Comparison of classification results with MLP in the south of the basin area. (a) RG mean+stdev, (b) MR mean+stdev and (c) MS mean+stdev
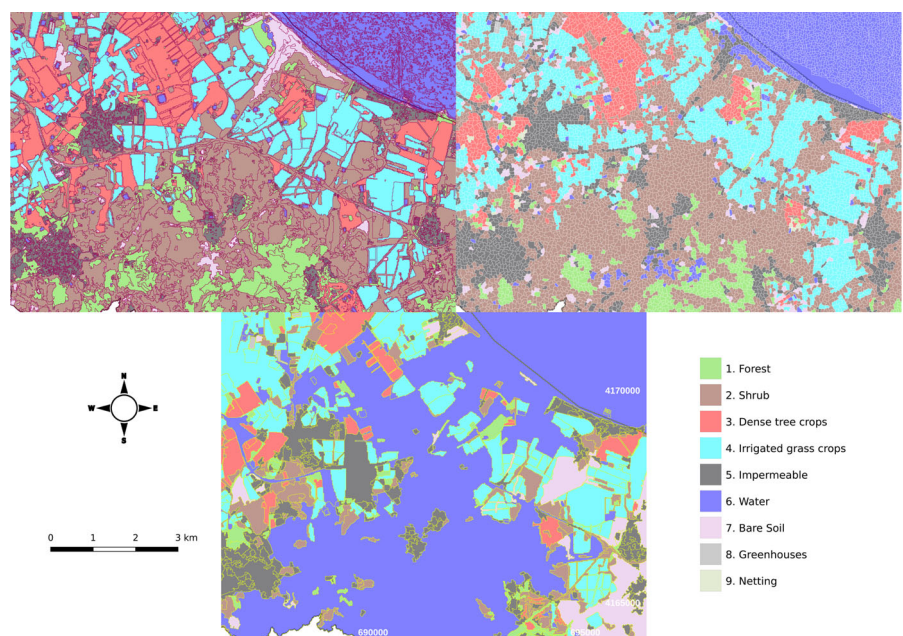
Many methods of parameter optimisation for segmentation, as well as evaluation of the results, have been proposed, with these aspects still remaining as issues to be solved (Hossain and Chen 2019), the trial and error method is still the most widely used due to its reliability. However, it is time-consuming, so this study proposes the automation of both methods, interpolating the results and selecting the most appropriate combination of parameters. Being aware of the partial subjectivity of this selection and of the fact that the classification statistics would have been different if the parameters had been slightly varied, the aim was to obtain the segmentation parameters together with the evaluation metrics of segmentation at the same time and with the same resources required for trial and error type tests. These metrics were based on minimising the intra-segment variability and maximising the inter-segment variability. In this way, the interpolation procedure provided a much wider range of parameter values from which the best compromise solution could be selected.

Kucharczyk et al (2020) highlights as a good practice to reduce the dimensionality of the models with a selection of variables, being according to Ma et al. (2017) between 15 and 25 the recommended number for RF. The selection made in this study with the RF models manages to reduce the number of variables to this range or very close to it in several of the datasets, and at least in one of the datasets of each segmentation algorithm.

However, the final assessment of the results of each segmentation algorithm is given by the evaluation of the entire GEOBIA process, carried out using the metrics obtained in each of the classifications. As some authors have considered the kappa index and the global adjustment to be insufficient and even of little use for remote sensing applications (Liu et al. 2007; Pontius and Millones 2011; Stehman and Foody 2019; Foody 2020), it was decided to additionally analyse the metrics by class from the confusion matrix, following Foody (2020). This improved the initial position of RG between the segmentation methods. Although it did not produce the best overall accuracy with either of the two classification methods used, is indeed the one with more balanced accuracy metrics by class.

In terms of classification method, the best performance was obtained with RF, as it was able to produce models with better per-class results as well as global accuracies. This is probably due to the greater robustness of RF to small datasets and the sensitivity of MLP to sparse training data.

RF also produced the best results in a previous study using a multi-sensor and multi-temporal pixel-based classification in the same area (Valdivieso-Ros et al. 2023), similar to those obtained with the best dataset in this study by cross-validation, with an overall accuracy (OA) of $0.91 \pm 0.005$ ($\alpha = 0.05$) and a kappa (K) of $0.898 \pm 0.006$ ($\alpha = 0.05$). However, per class and qualitatively, the pixel-based approach was more accurate than the de GEOBIA approach, with the omission and commission errors of the majority of classes being less than 0.1. Only the netting class obtained higher errors, 0.25 and 0.27, but still lower than those obtained in this GEOBIA approach. Similarly, the results obtained with MLP in the pixel-based approach, an OA of $0.877 \pm 0.006$ ($\alpha = 0.05$) and a K of $0.859 \pm 0.006$ ($\alpha = 0.05$), are almost equal to those obtained with the most accurate dataset in this study.

The reason for this is probably related to the resolution of the images, which is insufficient to obtain a good segmentation in this particularly fragmented area. When the pixel-based validation is carried out on the three segmentations, a very high accuracy and kappa values are achieved as the undersegmented objects are classified as the test pixel class, but without a proper delineation of recognisable objects. In this way, the quantitative results were extremely accurate, in contrast to the qualitative results. Nevertheless, the confusion observed between some classes (greenhouses and netting or between them and crops) is easily explained and does not represent a significant difference in agronomic terms.

Considering all the above means, it was found that finding an algorithm that produces perfect segmentation is very difficult due to the complexity of implementing the algorithms and the lack of free, easy to use and powerful software and tools with which to implement them, in agreement with Hossain and Chen (2019). Nevertheless, the proposed method to obtain the compromise parameters for segmentation algorithms has facilitate the selection of a good combination consuming less time than the usual trial and error method.

However, it is not only the segmentation process that influences the accuracy metrics in a GEOBIA process, but also the features extracted from the segments to form the predictor datasets in classification. Problems with segmentation could be compensated by a good selection of these features. Blaschke et al. (2008) pointed out that the objects obtained from the segmentation are analysed on the basis of their spectral, textural and geometrical characteristics, among others. In this study, these characteristics were obtained by calculating shape metrics of the segments, in addition to centrality and a dispersion statistic for the values of the variables within each segment. As the variables were multi-sensor, including textural and contextual variables derived from optical, SAR and LiDAR data, the overall classification approach was very complete. The improvement in the use of multi-sensor variables over results obtained using only one source of data, either optical or SAR or derived variables, in the same area has been demonstrated using a pixel-based approach (Valdivieso-Ros et al. 2023). There are little research in the same area using an object-based approach, although the doctoral thesis carried out by Fructuoso (2015) performed several classifications using optical data and some

segmentation techniques achieving OA results around 0.83 and K of 0.81. Furthermore, the results are consistent with those obtained using GEOBIA with multitemporal and multisensor variables in other areas, such as those carried out by Ruiz et al. (2021), Maxwell et al. (2019), Gbodjo et al. (2020) or Wu et al. (2016), with OA of more than 0.90. However, the use of multisensor data for GEOBIA classification is an aspect that, according to Johnson and Ma (2020), has not yet been well researched.

## Conclusions

The optimisation procedure has facilitated the selection of parameters from a wider range of values. Although not entirely based on trial and error, it is still a partially subjective selection, as the parameters are estimated by interpolation and the user chooses a compromise solution from the results; it cannot be excluded that some of the problems encountered in the subsequent segmentation would not have arisen with the choice of other parameters.

The segmentation algorithm that produces the best global metrics, the MR algorithm, is not effective in classifying some of the classes with either of the two algorithms used for classification. However, the second segmentation algorithm based on global accuracy metrics, RG, is better at classifying some difficult classes. In fact, this problem is partly due to a small training dataset. However, in a previous study by Valdivieso-Ros et al. (2023), which performed pixel

classification in the same area and with the same dataset, the global metrics were similar and the results did not show the misclassification of the netting class when using RF. On the other hand, RF has shown a better performance compared to the other classification method tested on the same data, MLP, which confirms that it is the best method to classify the area for the data used. Moreover, the proposed method of variable selection has proved its validity, in particular reducing the number of variables from 262 to a range of 15-27 among the best performing datasets.

The use of multi-sensor data increases the accuracy metrics compared to the use of a single data source when using a pixel-based approach, as appears to be the case with GEOBIA, although more research is needed on this topic.

Another important conclusion is the importance of using not only global accuracy metrics, but also class metrics in the evaluation of GEOBIA.

Some aspects remain unexplored and constitute future lines of research, such as the exploration of automatic methods for the selection of training segments, which could increase the number of samples to solve the problems of the most difficult classes. It would also be interesting to extend the variety of segmentation algorithms to determine whether other types, such as edge-based, watershed or neural network based models would be more effective in this area.

## Appendix A: Classification performance by class

**Table 8** RF accuracy metrics per class of each dataset and segmentation method

| Metrics | Data | Forest | Scrub | Dense tree crops | Irrigated crops | Impermeable | Water | Bare soil | Greenhouses | Netting |
|---------|------|--------|-------|------------------|-----------------|-------------|-------|-----------|-------------|---------|
| RG Precision | mean+stdev | 0.94 | 0.864 | 0.902 | 0.907 | 0.923 | 0.870 | 0.860 | 0.686 | 0.556 |
| | median+range | 0.974 | 0.809 | 0.900 | 0.886 | 0.923 | 0.889 | 0.848 | 0.767 | 0.667 |
| RG Recall | mean+stdev | 0.905 | 0.950 | 0.881 | 0.975 | 0.818 | 0.976 | 0.902 | 0.800 | 0.263 |
| | median+range | 0.905 | 0.950 | 0.857 | 0.975 | 0.818 | 0.976 | 0.951 | 0.767 | 0.316 |
| RG Balanced Accuracy | mean+stdev | 0.951 | 0.965 | 0.934 | **0.981** | 0.904 | 0.978 | 0.941 | 0.882 | 0.625 |
| | median+range | 0.951 | 0.960 | 0.922 | 0.979 | 0.904 | 0.979 | **0.964** | 0.872 | **0.653** |
| MS Precision | mean+stdev | 0.929 | 0.706 | 0.780 | 0.796 | 0.917 | 0.920 | 0.739 | 0.688 | 0.600 |
| | median+range | 0.964 | 0.833 | 0.854 | 0.833 | 0.944 | 0.958 | 0.760 | 0.774 | 0.750 |
| MS Recall | mean+stdev | 0.963 | 0.774 | 0.800 | 0.951 | 0.868 | 0.958 | 0.654 | 0.759 | 0.176 |
| | median+range | 1.000 | 0.968 | 0.875 | 0.976 | 0.895 | 0.958 | 0.731 | 0.828 | 0.176 |
| MS Balanced Accuracy | mean+stdev | 0.977 | 0.866 | 0.881 | 0.954 | 0.928 | 0.975 | 0.815 | 0.859 | 0.584 |
| | median+range | **0.998** | **0.971** | 0.925 | 0.971 | 0.943 | 0.977 | 0.853 | 0.899 | 0.586 |
| MR Precision | mean+stdev | 0.960 | 0.877 | 0.897 | 0.897 | 0.944 | 0.953 | 0.944 | 0.765 | 1.000 |
| | median+range | 0.958 | 0.836 | 0.881 | 0.897 | 0.895 | 1.000 | 0.875 | 0.794 | – |

**Table 8** continued

| Metrics | Data | Forest | Scrub | Dense tree crops | Irrigated crops | Impermeable | Water | Bare soil | Greenhouses | Netting |
|---|---|---|---|---|---|---|---|---|---|---|
| MR Recall | mean+stdev | 1.000 | 0.943 | 0.912 | 0.897 | 0.944 | 1.000 | 0.850 | 0.897 | 0.077 |
| | median+range | 0.958 | 0.962 | 0.912 | 0.897 | 0.944 | 1.000 | 0.700 | 0.931 | 0.000 |
| MR Balanced Accuracy | mean+stdev | 0.997 | 0.959 | **0.945** | 0.942 | **0.969** | **0.997** | 0.923 | 0.935 | 0.538 |
| | median+range | 0.976 | 0.963 | 0.944 | 0.942 | 0.966 | 1.000 | 0.847 | **0.954** | 0.500 |

LOPOCV type validation. The dataset with the highest accuracy per algorithm is highlighted in bold

**Table 9** RF omission and commission errors by class for each dataset and best-accuracy segmentation method with RF for each segmentation algorithm: RG mean+stdev, MR mean+stdev and MS median+range

| | | Forest | Scrub | Dense tree crops | Irrigated crops | Impermeable | Water | Bare soil | Greenhouses | Netting |
|---|---|---|---|---|---|---|---|---|---|---|
| RG | Err.Om. | 0.026 | 0.136 | 0.098 | 0.093 | 0.077 | 0.130 | 0.140 | 0.314 | 0.444 |
| | Err.Com. | 0.095 | 0.050 | 0.119 | 0.025 | 0.182 | 0.024 | 0.098 | 0.200 | 0.737 |
| MS | **Err.Om.** | 0.036 | 0.167 | 0.146 | 0.167 | 0.056 | 0.042 | 0.240 | 0.226 | 0.250 |
| | Err.Com. | 0.000 | 0.032 | 0.125 | 0.024 | 0.105 | 0.042 | 0.269 | 0.172 | 0.824 |
| MR | Err.Om. | 0.04 | 0.123 | 0.103 | 0.103 | 0.056 | 0.047 | 0.056 | 0.235 | 0.000 |
| | Err.Com. | 0.00 | 0.057 | 0.088 | 0.103 | 0.056 | 0.000 | 0.150 | 0.103 | 0.923 |

**Table 10** MLP accuracy metrics per class of each dataset and segmentation method. CV5F type validation

| | Estatistic | Data | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RG | Precision | mean+stdev | 1.000 | 0.774 | 0.850 | 0.951 | 0.895 | 0.958 | 0.500 | 0.828 | 0.000 |
| | | median+range | 0.952 | 0.850 | 0.905 | 0.975 | 0.841 | 0.951 | 0.780 | 0.833 | 0.053 |
| | Recall | mean+stdev | 0.931 | 0.615 | 0.850 | 0.830 | 0.895 | 0.920 | 0.619 | 0.750 | 0.000 |
| | | median+range | 0.930 | 0.791 | 0.905 | 0.886 | 0.860 | 0.886 | 0.821 | 0.658 | 0.333 |
| | Balanced Accuracy | mean+stdev | 0.966 | 0.649 | 0.675 | 0.815 | 0.697 | 0.793 | 0.500 | 0.683 | 0.053 |
| | | median+range | 0.765 | 0.695 | 0.702 | **0.860** | 0.661 | 0.800 | 0.629 | 0.690 | 0.217 |
| MS | Precision | mean+stdev | 1 | 0.871 | 0.875 | 0.951 | 0.921 | 0.958 | 0.692 | 0.862 | 0.118 |
| | | median+range | 1.000 | 0.839 | 0.900 | 0.951 | 0.895 | 1.00 | 0.577 | 0.862 | 0.235 |
| | Recall | mean+stdev | 1 | 0.818 | 0.875 | 0.848 | 0.921 | 0.958 | 0.692 | 0.735 | 0.400 |
| | | median+range | 0.964 | 0.684 | 0.923 | 0.848 | 0.919 | 0.96 | 0.682 | 0.758 | 0.800 |
| | Balanced Accuracy | mean+stdev | – | 0.709 | 0.688 | 0.813 | 0.711 | 0.729 | 0.596 | **0.714** | 0.283 |
| | | median+range | 0.982 | 0.695 | 0.676 | 0.813 | 0.674 | **0.98** | 0.535 | 0.712 | **0.436** |
| MR | Precision | mean+stdev | 1.000 | 0.943 | 0.912 | 0.872 | 0.917 | 0.976 | 0.850 | 0.862 | 0.154 |
| | | median+range | 0.979 | 0.943 | 0.895 | 0.846 | 0.833 | 1 | 0.800 | 0.828 | 0.154 |
| | Recall | mean+stdev | 0.941 | 0.862 | 0.881 | 0.895 | 0.868 | 1.000 | 0.944 | 0.806 | 0.667 |
| | | median+range | 0.959 | 0.847 | 0.879 | 0.825 | 0.938 | 1 | 0.842 | 0.800 | 0.250 |
| | Balanced Accuracy | mean+stdev | **0.971** | 0.795 | **0.732** | 0.670 | **0.747** | 0.500 | 0.597 | 0.703 | 0.375 |
| | | median+range | 0.813 | **0.799** | 0.709 | 0.682 | 0.594 | – | **0.635** | 0.673 | 0.301 |

The dataset with the highest Balanced Accuracy by class is highlighted in bold

**Table 11** MLP accuracy metrics per class of each dataset and segmentation method. CV5F type validation

|    |         | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     |
|----|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| RG | Err.Om. | 0.071 | 0.275 | 0.071 | 0.075 | 0.159 | 0.049 | 0.195 | 0.167 | 1     |
|    | Err.Com.| 0.114 | 0.237 | 0.152 | 0.140 | 0.075 | 0.093 | 0.283 | 0.342 | 1     |
| MS | Err.Om. | 0     | 0.129 | 0.125 | 0.049 | 0.079 | 0.042 | 0.308 | 0.138 | 0.882 |
|    | Err.Com.| 0     | 0.182 | 0.125 | 0.152 | 0.079 | 0.042 | 0.308 | 0.265 | 0.600 |
| MR | Err.Om. | 0.000 | 0.057 | 0.088 | 0.128 | 0.083 | 0.024 | 0.150 | 0.138 | 0.846 |
|    | Err.Com.| 0.059 | 0.138 | 0.119 | 0.105 | 0.132 | 0.000 | 0.056 | 0.194 | 0.333 |

## Declarations

## References

Abadi M, Agarwal A, Barham P et al (2015) TensorFlow: Large-scale machine learning on heterogeneous systems. URL https://www.tensorflow.org/, software available from tensorflow.org

Adams R, Bischof L (1994) Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence 16:641–647. https://doi.org/10.1109/34.295913

Amoakoh AO, Aplin P, Awuah KT et al (2021) Testing the contribution of multi-source remote sensing features for random forest classification of the greater amanzule tropical peatland. Sensors 21:3399. https://doi.org/10.3390/s21103399

Anjna E, Kaur R (2017) Review of image segmentation techniques. International J Adva Res Comput Sci 8:36–39. https://doi.org/10.26483/ijarcs.v8i4.3691

Baatz M, Schape A (2000) Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: XII Engewandte Geographische Informationsverarbeitung. Wichmann Verlag

Belgiu M, Drăgu L (2016) Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing 114:24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Blaschke T, Lang S, Hay G (2008) Object-based image analysis. Springer, Berlin Heidelberg,. https://doi.org/10.1007/978-3-540-77058-9

Blaschke T, Hay GJ, Kelly M et al (2014) Geographic object-based image analysis - towards a new paradigm. ISPRS J Photogramm and Remote Sens 87:180–191. https://doi.org/10.1016/j.isprsjprs.2013.09.014

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Cánovas-García F, Alonso-Sarría F (2015) Optimal combination of classification algorithms and feature ranking methods for object-based classification of submeter resolution z/i-imaging dmc imagery. Remote Sens 7:4651–4677. https://doi.org/10.3390/rs70404651

Cánovas-García F, Alonso-Sarría F, Gomariz-Castillo F et al (2017) Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery. Computers and Geosciences 103:1–11. https://doi.org/10.1016/j.cageo.2017.02.012

CARM (2023) Estadística agraria regional. "figshare https://www.carm.es/web/pagina?IDCONTENIDO=1174?&RASTRO=c1415M&IDTIPO=100

Chen L, Ren C, Bao G et al (2022) Improved object-based estimation of forest aboveground biomass by integrating lidar data from gedi and icesat-2 with multi-sensor images in a heterogeneous mountainous region. Remote Sensing 14:2743. https://doi.org/10.3390/rs14122743

Câmara G, Vinhas L, Ferreira KR, et al (2008) Terralib: An open source gis library for large-scale environmental and socio-economic applications. In: Hall GB, Leahy M (eds) Open source approaches in spatial data handling. Advances in Geographic Information Science. Springer, Berlin, Heidelberg. pp 247–270. https://doi.org/10.1007/978-3-540-74831-1_12

Cnovas-Garca F, Alonso-Sarra F (2015) A local approach to optimize the scale parameter in multiresolution segmentation for multispectral imagery. Geocarto International 30:937–961. https://doi.org/10.1080/10106049.2015.1004131

Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24:603–619. https://doi.org/10.1109/34.1000236

Dobrinić D, Medak D, Gašparović M (2020) Integration of multi-temporal sentinel-1 and sentinel-2 imagery for land-cover classification using machine learning methods. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B1-2020:91–98. https://doi.org/10.5194/isprs-archives-XLIII-B1-2020-91-2020

Foody GM (2020) Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. Remote Sens Environ 239:111630. https://doi.org/10.1016/j.rse.2019.111630

Fructuoso MFC (2015) Seguimiento de los cambios de usos y su influencia en las comunidades y hhbitats naturales en la cuenca del mar menor, 1988-2009, con el uso de sig y teledeteccin. PhD thesis, Faculty of Biology. University of Murcia

Fukunaga K, Hostetler L (1975) The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Transactions on Information Theory 21:32–40. https://doi.org/10.1109/TIT.1975.1055330

Gao Y, Mas JF (2008) A comparison of the performance of pixel-based and object-based classifications over images with various spatial resolutions. Online J Earth Sci 2(1):27–35

Gbodjo YJE, Ienco D, Leroux L, et al (2020) Object-based multi-temporal and multi-source land cover mapping leveraging hierarchical class relationships. Remote Sensing 12(17). https://doi.org/10.3390/rs12172814, URL https://www.mdpi.com/2072-4292/12/17/2814

Georganos S, Grippa T, Lennert M et al (2018) Scale matters: Spatially partitioned unsupervised segmentation parameter optimization for large and heterogeneous satellite images. Remote Sens 10:1440. https://doi.org/10.3390/rs10091440

Giménez-Casalduero F, Gomariz-Castillo F, Alonso-Sarría F et al (2020) Pinna nobilis in the mar menor coastal lagoon: a story of colonization and uncertainty. Marine Ecology Progress Series 652:77–94. https://doi.org/10.3354/meps13468

Gomariz-Castillo F, Alonso-Sarría F, Cánovas-García F (2017) Improving classification accuracy of multi-temporal landsat images by assessing the use of different algorithms, textural and ancillary information for a mediterranean semiarid area from 2000 to 2015. Remote Sens 9:1058. https://doi.org/10.3390/rs9101058

Gonçalves J, Pôças I, Marcos B et al (2019) Segoptim-a new r package for optimizing object-based image analyses of high-spatial resolution remotely-sensed data. Int J Appl Earth Obs Geoinformation 76:218–230. https://doi.org/10.1016/j.jag.2018.11.011

Grippa T, Georganos S, Vanhuysse SG et al (2017) A local segmentation parameter optimization approach for mapping heterogeneous urban environments using vhr imagery. In: Heldens W, Chrysoulakis N, Erbertseder T, et al (eds) Proceedings Volume 10431, Remote Sensing Technologies and Applications in Urban Environments II. SPIE, pp 20, https://doi.org/10.1117/12.2278422

Hahnloser RHR, Sarpeshkar R, Mahowald MA et al (2000) Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature 405:947–951. https://doi.org/10.1038/35016072

Hanbury A (2009) Image segmentation by region based and watershed algorithms. In: Wah BW (ed) Wiley encyclopedia of computer science and engineering. John Wiley and Sons, Inc., Hoboken, NJ, https://doi.org/10.1002/9780470050118.ecse614

Hossain MD, Chen D (2019) Segmentation for object-based image analysis (obia): A review of algorithms and challenges from remote sensing perspective. ISPRS Journal of Photogrammetry and Remote Sensing 150:115–134. https://doi.org/10.1016/j.isprsjprs.2019.02.009

James G, Witten D, Hastie T et al (2013) An Introduction to Statistical Learning, vol 103. Springer, New York,. https://doi.org/10.1007/978-1-4614-7138-7

Johnson BA, Ma L (2020) Image segmentation and object-based image analysis for environmental monitoring: Recent areas of interest, researchers views on the future priorities. Remote Sensing 12:1772. https://doi.org/10.3390/rs12111772

Jozdani SE, Johnson BA, Chen D (2019) Comparing deep neural networks, ensemble classifiers, and support vector machine algorithms for object-based urban land use/land cover classification. Remote Sensing 11(14). https://doi.org/10.3390/rs11141713

Kingma DP, Ba J (2015) Adam: A method for stochastic optimization. URL http://arxiv.org/abs/1412.6980, paper presented at the International Conference on Learning Representations (ICLR), San Diego, CA, 7–9 May 2015

Kucharczyk M, Hay GJ, Ghaffarian S et al (2020) Geographic object-based image analysis: A primer and future directions. Remote Sensing 12:2012. https://doi.org/10.3390/rs12122012

Lang S (2008) Object-based image analysis for remote sensing applications: modeling reality - dealing with complexity. In: Broy M, Denert E (eds) Object-based image analysis. Springer, Berlin, Heidelberg, pp 3–27. https://doi.org/10.1007/978-3-540-77058-9_1

Liaw A, Wiener M (2002) Classification and regression by randomforest. R News 2:18–22

Liaw A, Yan J, Li W et al (2014) Classification and regression by randomforest. R news XXXIX

Liu C, Frazier P, Kumar L (2007) Comparative assessment of the measures of thematic classification accuracy. Remote Sensing of Environment 107:606–616. https://doi.org/10.1016/j.rse.2006.10.010

Liu S, Qi Z, Li X et al (2019) Integration of convolutional neural networks and object-based post-classification refinement for land use and land cover mapping with optical and sar data. Remote Sensing 11:690. https://doi.org/10.3390/rs11060690

Luca GD, Silva JMN, Fazio SD et al (2022) Integrated use of sentinel-1 and sentinel-2 data and open-source machine learning algorithms for land cover mapping in a mediterranean region. European Journal of Remote Sensing 55:52–70. https://doi.org/10.1080/22797254.2021.2018667

Ma L, Fu T, Blaschke T et al (2017) Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers. ISPRS International Journal of Geo-Information 6:51. https://doi.org/10.3390/ijgi6020051

Ma L, Liu Y, Zhang X et al (2019) Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J Photogramm Remote Sens 152:166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Martínez J, Esteve M, Martínez-Paz J et al (2007) Simulating management options and scenarios to control nutrient load to mar menor, southeast spain. Transitional Waters Monographs TWM, Transit Waters Monogr 1:53–70. https://doi.org/10.1285/i18252273v1n1p53

Masiza W, Chirima JG, Hamandawana H et al (2020) Enhanced mapping of a smallholder crop farming landscape through image fusion

and model stacking. Int J Remote Sens 41:8739–8756. https://doi.org/10.1080/01431161.2020.1783017

Maxwell AE, Warner TA, Fang F (2018) Implementation of machine-learning classification in remote sensing: an applied review. International Journal of Remote Sensing 39:2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Maxwell AE, Strager MP, Warner TA, et al (2019) Large-area, high spatial resolution land cover mapping using random forests, geobia, and naip orthophotography: Findings and recommendations. Remote Sensing 11(12). https://doi.org/10.3390/rs11121409, URL https://www.mdpi.com/2072-4292/11/12/1409

McInerney D, Kempeneers P (2015) Open Source Geospatial Tools. Springer International Publishing. https://doi.org/10.1007/978-3-319-01824-9

Mitasova H, Neteler M (2008) Open source GIS: a GRASS GIS approach. Springer

Pontius RG, Millones M (2011) Death to kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. Int J Remote Sens 32:4407–4429. https://doi.org/10.1080/01431161.2011.552923

Qin R, Liu T (2022) A review of landcover classification with very-high resolution remotely sensed optical images-analysis unit, model scalability and transferability. Remote Sensing 14:646. https://doi.org/10.3390/rs14030646

Reichstein M, Camps-Valls G, Stevens B et al (2019) Deep learning and process understanding for data-driven earth system science. Nature 566:195–204. https://doi.org/10.1038/s41586-019-0912-1

Rodriguez-Galiano V, Ghimire B, Rogan J et al (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 67:93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Räsänen A, Rusanen A, Kuitunen M et al (2013) What makes segmentation good? a case study in boreal forest habitat mapping. International Journal of Remote Sensing 34:8603–8627. https://doi.org/10.1080/01431161.2013.845318

Ruiz LFC, Guasselli LA, Simioni JPD et al (2021) Object-based classification of vegetation species in a subtropical wetland using sentinel-1 and sentinel-2a images. Sci Remote Sens 3:100017. https://doi.org/10.1016/j.srs.2021.100017

Shepherd J, Bunting P, Dymond J (2019) Operational large-scale segmentation of imagery based on iterative elimination. Remote Sens 11:658. https://doi.org/10.3390/rs11060658

Šiljeg A, Panda L, Domazetović F, et al (2022) Comparative assessment of pixel and object-based approaches for mapping of olive tree crowns based on uav multispectral imagery. Remote Sensing 14(3). https://doi.org/10.3390/rs14030757

Stehman SV, Foody GM (2019) Key issues in rigorous accuracy assessment of land cover products. Remote Sensing of Environment 231:111199. https://doi.org/10.1016/j.rse.2019.05.018

Team RC (2020) R: A language and environment for statistical computing. URL https://www.R-project.org/index.html

Valdivieso-Ros C, Alonso-Sarria F, Gomariz-Castillo F (2023) Effect of the synergetic use of sentinel-1, sentinel-2, lidar and derived data in land cover classification of a semiarid mediterranean area using machine learning algorithms. Remote Sensing 15:312. https://doi.org/10.3390/rs15020312

Wang J, Bretz M, Dewan MAA et al (2022) Machine learning in modelling land-use and land cover-change (lulcc): Current status, challenges and prospects. Sci Total Environ 822:153559. https://doi.org/10.1016/j.scitotenv.2022.153559

Whiteside TG, Boggs GS, Maier SW (2011) Comparing object-based and pixel-based classifications for mapping savannas. International Journal of Applied Earth Observation and Geoinformation 13(6):884–893. https://doi.org/10.1016/j.jag.2011.06.008

Wu M, Sun Z, Yang B et al (2016) A hierarchical object-oriented urban land cover classification using worldview-2 imagery and airborne lidar data. In: IOP Conference Series: Earth and Environmental Science, IOP Publishing, pp 012016, https://doi.org/10.1088/1755-1315/46/1/012016

Wurm M, Stark T, Zhu XX et al (2019) Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. ISPRS J Photogramm Remote Sens 150:59–69. https://doi.org/10.1016/j.isprsjprs.2019.02.006