ORIGINAL ARTICLE

Open Access

CrossMark

Open source R for applying machine learning to RPAS remote sensing images

Marco Piragnolo[*], Andrea Masiero and Francesco Pirotti

Abstract

The increase in the number of remote sensing platforms, ranging from satellites to close-range Remotely Piloted Aircraft System (RPAS), is leading to a growing demand for new image processing and classification tools. This article presents a comparison of the Random Forest (RF) and Support Vector Machine (SVM) machine-learning algorithms for extracting land-use classes in RPAS-derived orthomosaic using open source R packages.
The camera used in this work captures the reflectance of the Red, Blue, Green and Near Infrared channels of a target. The full dataset is therefore a 4-channel raster image. The classification performance of the two methods is tested at varying sizes of training sets. The SVM and RF are evaluated using Kappa index, classification accuracy and classification error as accuracy metrics. The training sets are randomly obtained as subset of 2 to 20% of the total number of raster cells, with stratified sampling according to the land-use classes. Ten runs are done for each training set to calculate the variance in results. The control dataset consists of an independent classification obtained by photointerpretation. The validation is carried out(i) using the K-Fold cross validation, (ii) using the pixels from the validation test set, and (iii) using the pixels from the full test set.
Validation with K-fold and with the validation dataset show SVM give better results, but RF prove to be more performing when training size is larger. Classification error and classification accuracy follow the trend of Kappa index.

Keywords: Remote sensing, R software, Machine learning, Random forest, Support vector machine, RPAS, Land use classification

Background

The increase in the number of remote sensing platforms, ranging from satellites to close-range Remotely Piloted Aircraft System (RPAS), is causing a growing demand for new tools for image processing and classification. Classification is applied in many research fields such as geomorphology, environmental analyses, land use, fragmentation of habitats and risk assessment [1, 2] just to name a few. In particular RPAS are applied to fields that benefit from close-range sensing, such as 3D modelling of cultural heritage and archaeology, environmental sciences, precision forestry and precision agriculture [3–7]. Imagery collected from remote sensing platforms is commonly classified using conventional remote sensing techniques supplied by available software in the market. In remote sensing literature, there are two main classification approaches, pixel-based and object-based. The pixel-based methods can be divided into unsupervised and supervised. The unsupervised classifiers cluster pixels in a number of classes based on statistical information from the image. The process is automatic and the user can only set the number of clusters. The supervised classifiers are based on training areas inserted by an operator, which define a spectral signature for each class. The object-based classifiers defined an object using geometric information, contextual information and texture information.

Machine learning techniques are classification/regression methods for analysing data. They can be used for supervised and unsupervised classification. They use algorithms that learn from previous computation, and they were recently applied in investigations regarding cotton crop [8], variable-rate fertilization [9], classification of invasive weed species [10], detecting landing sites [11, 12], geological mapping [13], Land Use/Land Cover (LULC) classification [14–18].

* Correspondence: marco.piragnolo@phd.unipd.it
CIRGEO, Interdepartmental Research Center of Geomatics, University of Padua, Viale dell'Università 16, 35020 Legnaro, Italy

Springer Open

Piragnolo *et al. Open Geospatial Data, Software and Standards* (2017) 2:16

Page 2 of 7

Recent developments in technology have pushed for a fast increase in using RPAS – commonly referred to as "drones" for observation of the earth surface. The challenge of processing imagery obtained from RPAS resides in the increase of the size of datasets, which is due to increasing resolution of images and the ability of RPASs to collect hundreds of images in each flight. A novel approach using machine learning might provide faster and more accurate results than typical supervised classification of such images. The goal of this work is to benchmark the performance of two machine learning algorithms for classifying an RPAS-derived orthomosaic using open source R packages. The algorithms are Random Forest (RF) and Support Vector Machine (SVM). They are evaluated using three accuracy metrics, Kappa index, classification accuracy and classification error.

Material and methods

The RPAS images have been acquired in a testing area inside the Campus of Agripolis at University of Padova in the city of Legnaro (Italy). The size of the area is 241 m × 508 m. It contains heterogeneous land-cover, including bare ground, vegetation and urban features. The ground-truth is defined by direct observation. Eighteen ground control points (GCP) were defined in the area for orientation of the photogrammetric image block. The coordinates were collected with GNSS in Real Time Kinematic mode; the root mean square error (RMSE) of measures resulted between 0.008 and 0.011.

The RPAS flight was performed in November 2015 using a camera with Red, Blue, Green and Near Infrared camera (RGBI) carried by the SenseFly EBee fixed wing platform. The average ground sampling distance (GSD) was 4.5 cm on the ground at a flight altitude of 150 m. The images have been processed using Agisoft Photoscan. The result is an ortho-rectified mosaic of images, with an RMSE of 0.393 pixel. The final GSD, or spatial resolution,

is 6 cm, so the final dimensions are 4020 X 8466 pixels, and the storage size is 48.9 MB. To reduce the computation time, the full dataset was resampled using the nearest neighbour algorithm to a cell size of 30 cm. The nearest neighbour algorithm preserves the radiometric values of cells. Then, the orthomosaic is clipped to a final dimension of 801 × 529 pixels, storage size of 1.21 MB (Fig 1). The RF and SVM machine learning methods were tested on the clipped image, using the R/rminer package [19] available in The Comprehensive R Archive Network repository [20].

The R/rminer package, version 1.4.1 for R is an aggregator of 14 classification and 15 regression methods. It also includes methods for determining common accuracy metrics over results [21, 22]. Two algorithms, Support Vector Machine (SVM) and Random Forest (RF) have been compared in this study. The SVM uses a separating hyperplane as a predictor. A decision plane divides dataset into two groups. Hence, the set of objects has different class memberships, and data are transformed in classes by using a mathematical function called kernel [23]. The RF classifier consists of a collection of trees. It samples randomly the original dataset, and defines decision trees using bootstrap aggregating. Bootstrap is a statistical technique that allows approximating statistics (e.g. average, variance, confidence interval) of data from the data itself. It is used when the distribution of the original dataset is not known beforehand. A complete tree with all branches is grown for each sample, and the predictors are applied to each branch [24]. Finally, the best variable obtained from the predictor is chosen, and predictions are aggregated in a new sample. Consequently, a new sample is predicted, and the estimation of errors can be calculated at the level of iteration and aggregation [25, 26]. In this study, RF and SVM have been trained using a subset of 2 to 20% of the total number of raster cells. For each

Fig. 1 Clipped testing area: true colour image (*left*), false colour infrared image (*right*)

Piragnolo *et al. Open Geospatial Data, Software and Standards* (2017) 2:16

Page 3 of 7

percentage, ten training sets were extracted using stratified random sampling. This allowed to assess the variance from accuracy results for each size of training set. The control dataset is an independent classification based on photo interpretation as shown in Fig. 2. The classes for LULC are: (i) broadleaf, (ii) building, (iii) grass, (iv) headland access path, (v) road, (vi) sowed land, (vii) vegetable.

The framework of the benchmarking process is illustrated in Fig. 3. Each class in the area is represented differently in terms of number of pixels (i.e. area). Therefore, the number of pixels we sampled for training was proportional to the class area (i.e. stratified sampling). Pixels falling across two polygons, thus mixing two different classes, were discarded to limit using pixels with mixed spectral signature. For each set of stratified samples, ten different training sets and ten validation sets have been created. The training set is used to fit the model and to apply it for classification of the image. The validation dataset is the difference between the full set and the training set.

The framework trains and tests each of the two methods (RF and SVM) fitting the model and applying K-fold cross-validation. The K-fold cross-validation technique splits the data in K (10) sets (folds) of equal size. K – 1 subsamples are used as training test set, and a single subsample is used for validation. The procedure is repeated K times, but each subset is used only once for the validation.

The accuracy metrics used for comparing results are the Kappa index, the classification accuracy and the classification error. Their values range from 0 to 100, and they are estimated with three different approaches, (i) using pixels from the training test set and applying K-Fold cross validation, (ii) using pixels from the validation test set, and (iii) using pixels from the full test set.
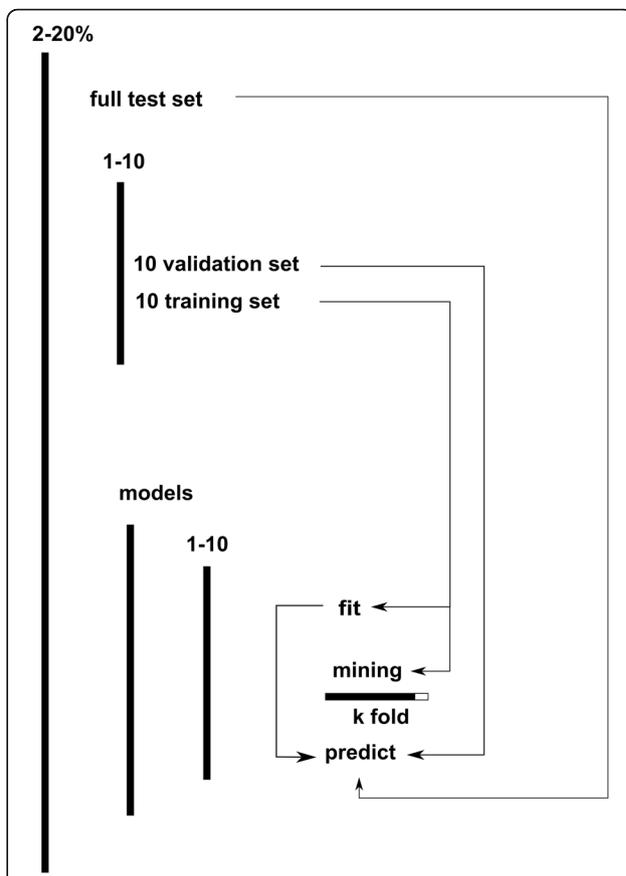

Fig. 3 The schema of the framework for benchmarking. Models (RF and SVM) have been applied to stratified samples ranging from 2 to 20% of the total population (n. of pixels). For each set of stratified samples, ten different training sets and ten validation sets have been extracted
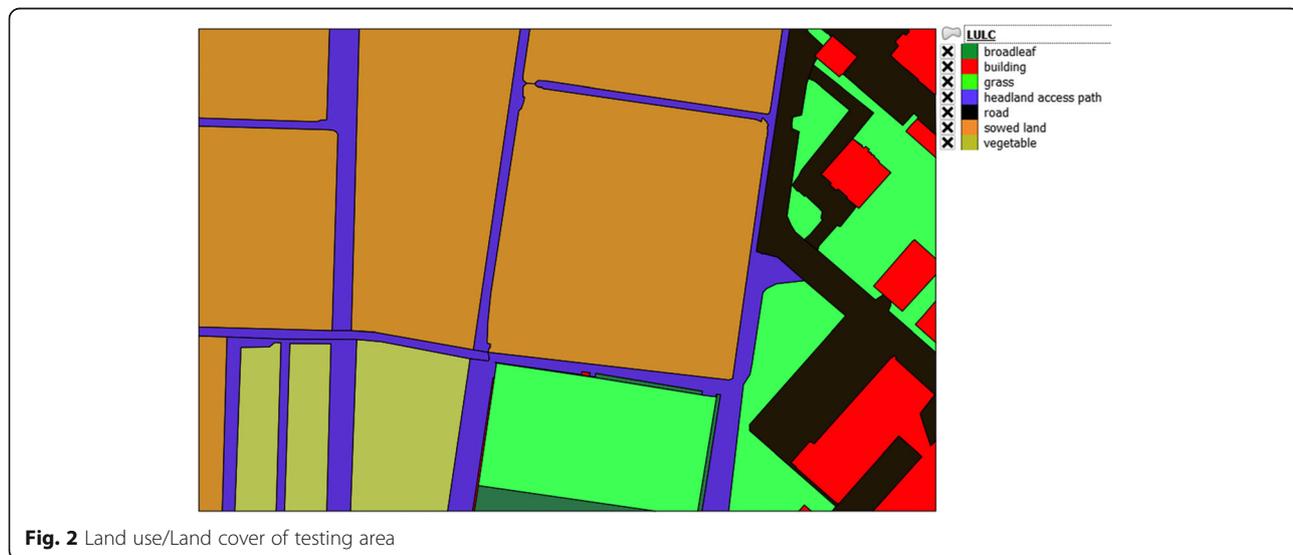

Fig. 2 Land use/Land cover of testing area

Piragnolo *et al. Open Geospatial Data, Software and Standards* (2017) 2:16
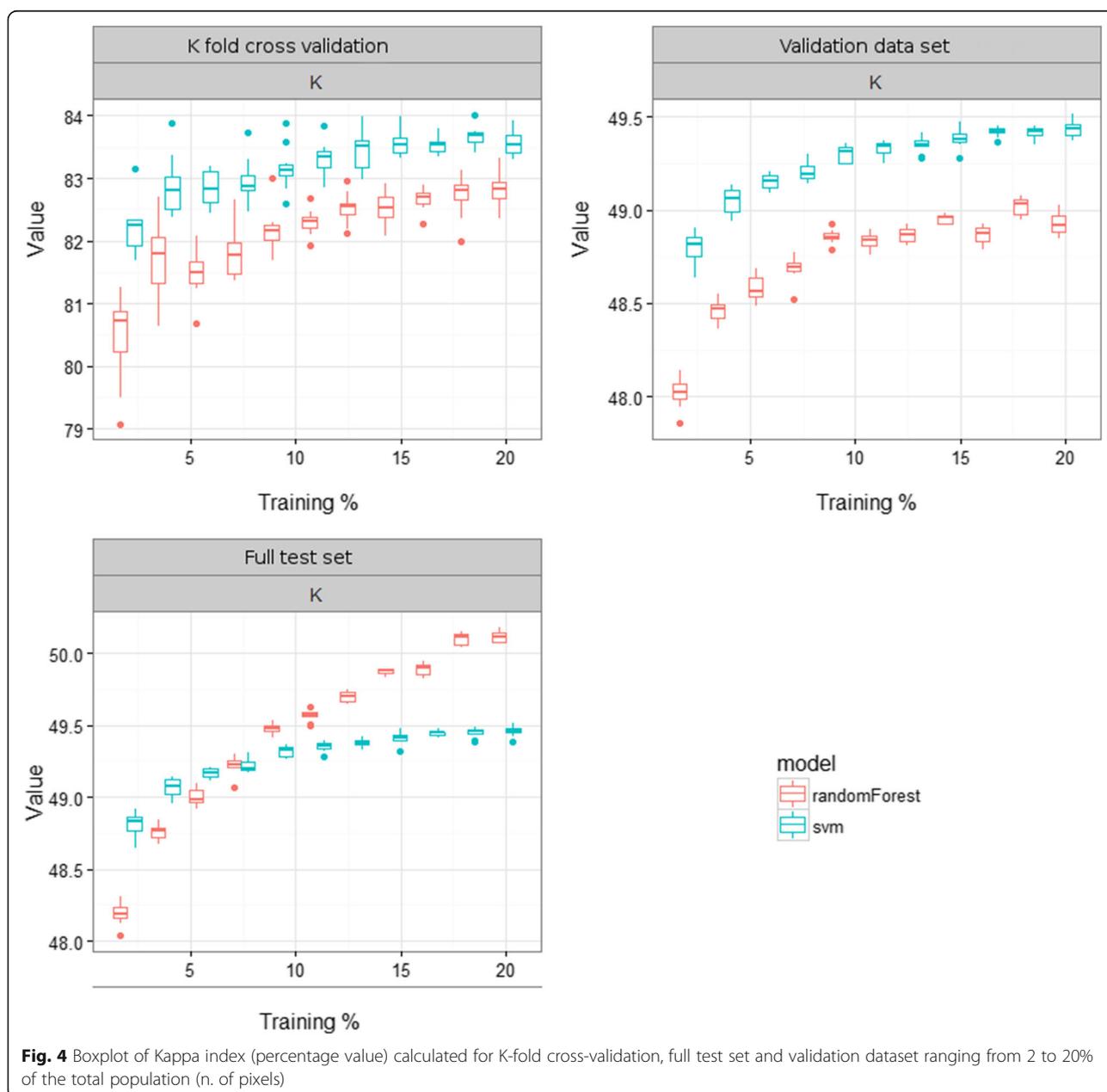
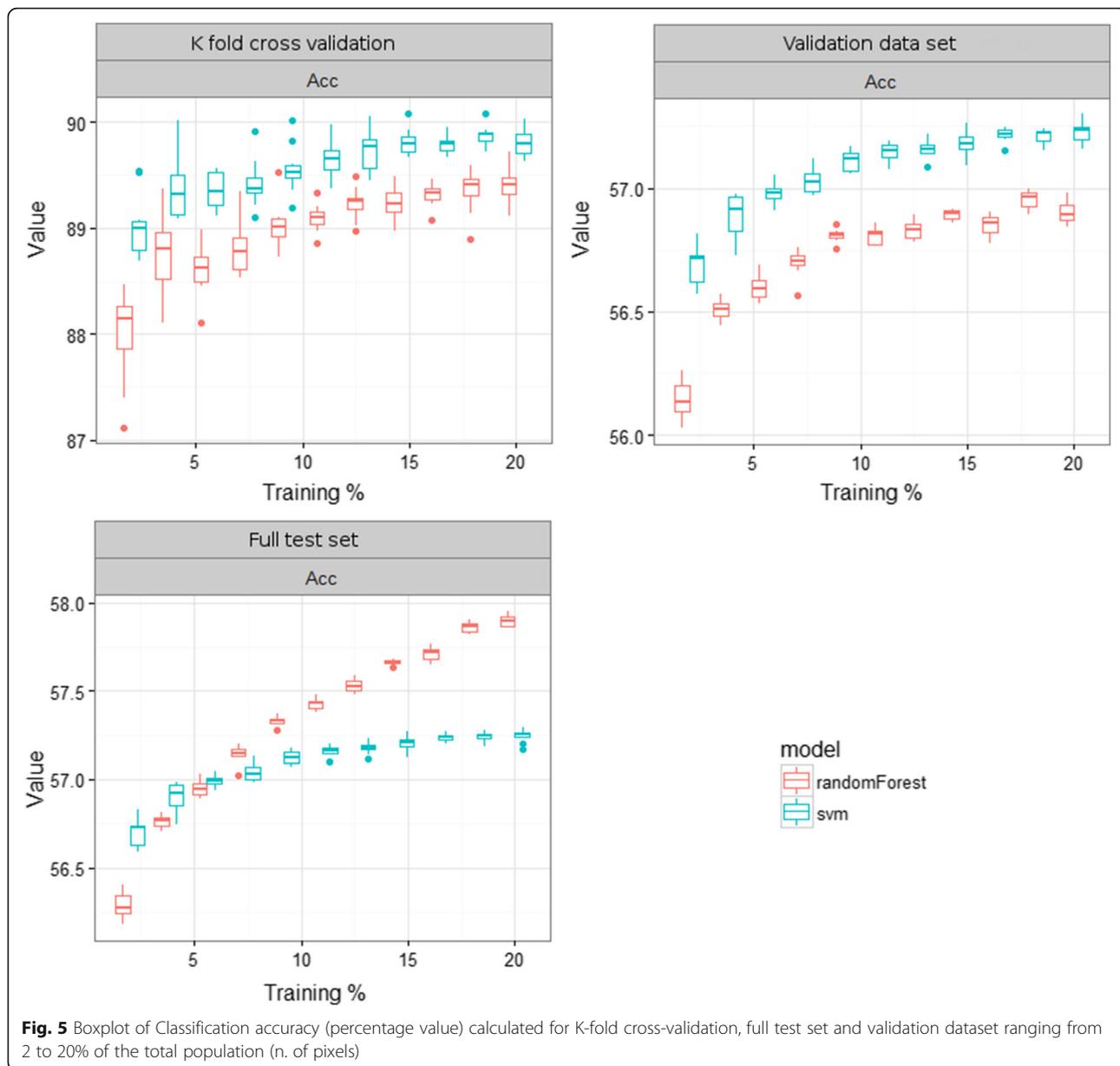Page 4 of 7

Results and discussion

The accuracy metrics are reported in three boxplots (Figs. 4, 5, 6) which represent respectively the Kappa index (K), the classification accuracy rate (Acc) and the classification error rate (CE). The last two are the inverse of each other. All metrics range from 0 to 100. The boxplots show the variance calculated from the ten runs for each training size.

Figure 4 reports Kappa index calculated for K-fold test set, validation test set and full test sets. Comparing the boxplots of the three validation methods, it is clear how their values grow proportionally with the training subset size from 2% to 20%. The variance decreases with the increase of the training subset size. In the K-fold cross validation test set, the results range from 80 to 84, and the SVM performs better than RF. Using the validation test set, the results are similar, but values are, as expected, lower, ranging from 48 to 49.5. In addition, in this case the SVM is better in comparison with RF. When validating against the full test set, the RF returns a better result than SVM when training with more than 10% of the full test set. The Kappa values range from 48 to 51. RF rises gradually from 48.5 to 50.5, whereas the SVM remains stable around 49.5.

Figure 5 shows the classification accuracy for K-fold test set, validation test set and full test sets. Likewise, the

Fig. 4 Boxplot of Kappa index (percentage value) calculated for K-fold cross-validation, full test set and validation dataset ranging from 2 to 20% of the total population (n. of pixels)

Piragnolo *et al. Open Geospatial Data, Software and Standards* (2017) 2:16

Page 5 of 7

Fig. 5 Boxplot of Classification accuracy (percentage value) calculated for K-fold cross-validation, full test set and validation dataset ranging from 2 to 20% of the total population (n. of pixels)

accuracy trend is similar the K index trend, and a gradual improvement in accuracy is related to an increasing training percentage. Indeed, using K-fold cross validation test set, the SVM gets a better result than RF. In this case, the score is over 88 whereas using a validation test the score ranges from 56 to 58. Using a full test set, the RF has better results than SVM, and the score ranges from 56 to 58. The RF rises gradually from 56 to 58, whereas SVM remains stable between 57 and 57.5.

Figure 6 illustrates a decreasing trend for classification error in the three dataset. In the K-fold cross validation test set, the SVM has less error that RF with variations of 0.5. Likewise, the validation test set has a decreasing trend, and the score ranges between 42 and 44. Using

the full test set, errors range from 42 to 43.5, but the RF and SVM have a similar score using less than 5% of the pixels as training. Using a set for training of more than 5% of the total pixels, the RF and SVM results have some differences. The RF has a decreasing trend, and it reaches the minimum around 42, whereas SVM remains steady at 42.7.

Results by no means intend to prove one classifier better than another. Classifiers behave differently depending on several factors, and results prove exactly this point. Figure 4 is the most informative, where the first two validation methods show a better performance by SVM, but validation against the full test set provides a different result. The comparison of results from three different validation
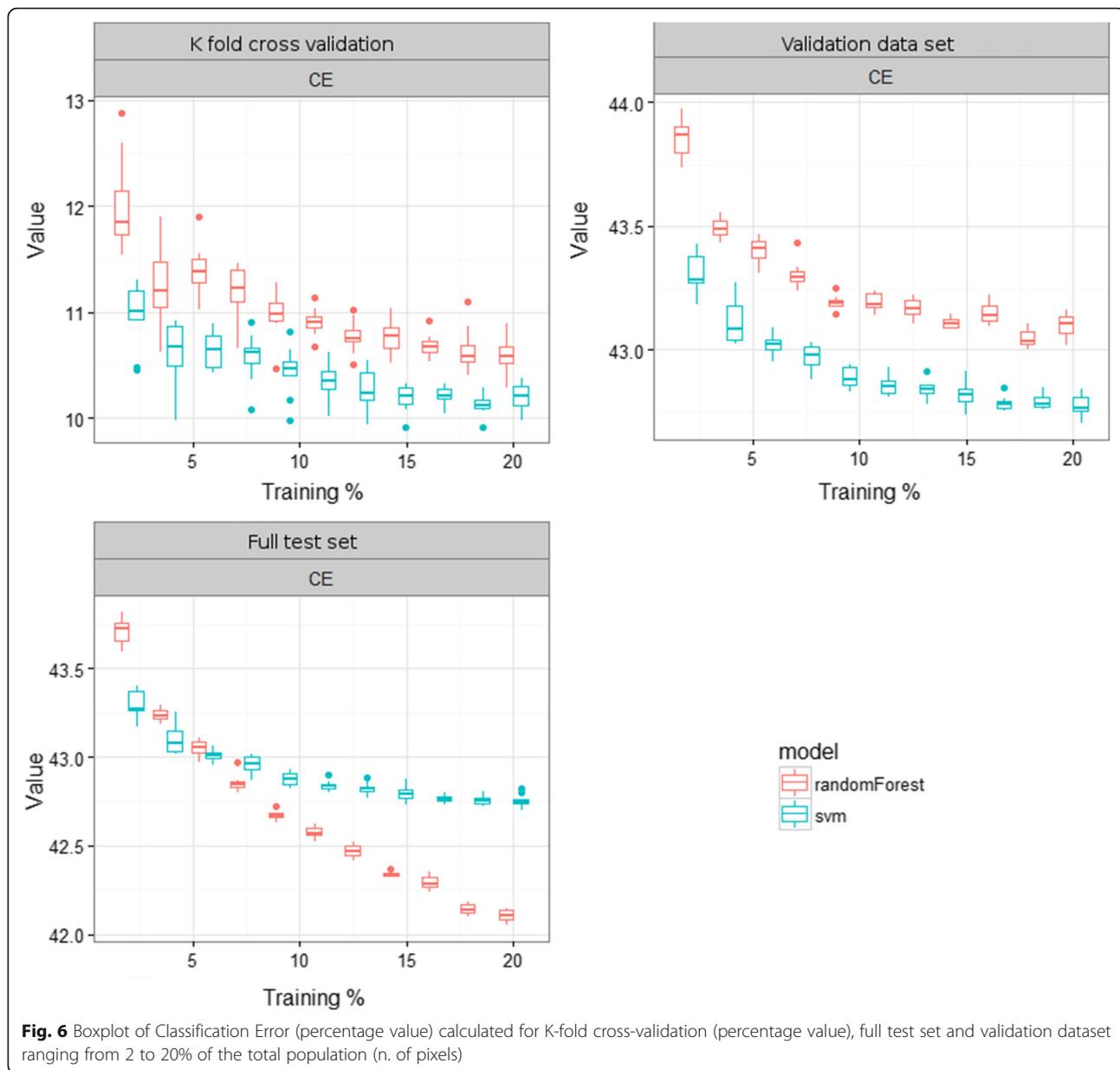
Piragnolo *et al. Open Geospatial Data, Software and Standards* (2017) 2:16

Page 6 of 7

Fig. 6 Boxplot of Classification Error (percentage value) calculated for K-fold cross-validation (percentage value), full test set and validation dataset ranging from 2 to 20% of the total population (n. of pixels)

methods provides added insights on the behaviour that operators can expect from the two classifiers. Also, another informative aspect from the plots is the added value from using larger training sets. This is an important aspect, as bigger training sets require more computing time, and relative expenditure in terms of energy. Knowing the range of improvement over a growing training size can support decision in future classification procedures.

Another source for discussion is the definition of classes and their identification over the image. This, of course, has a certain degree of subjectivity depending on the operator who manually defines these areas with polygons. Also, the inevitable aspect of inter-class and intra-class spectral mixtures has to be accounted for.

Border pixels were removed in this study to limit mixing, but this operation does not remove the problem completely. Nevertheless, the results show significant relative differences correlated to the size of training sets. This is something to consider as supportive information when using such methods.

Conclusions

This paper compared accuracy metrics of two machine learning algorithms, SVM and RF, using three validation methods and testing different sizes of training sets. As expected, accuracy was better when a bigger training size is used, but this trend is not linear. This is particularly evident when the validation is done against the full

Piragnolo *et al. Open Geospatial Data, Software and Standards* (2017) 2:16

Page 7 of 7

test set. SVM gets better results with smaller training sets, whereas RF becomes better at training sizes larger than 7–8% of the total. Validation with K-fold and with the validation dataset showed SVM give better results, but RF proved to be more performing when training size is larger. Classification error and classification accuracy followed the trend of Kappa index.

Future investigations will limit mixing by careful selection of single pixels for both training and validation. This will decrease the size of the sets, but will increase the purity of pixel class and provide better insight on the behaviour of the machine learning methods. Available multi-spectral imagery benchmarking datasets will be considered also for further testing, for example the MUULF Gulfport dataset [27]. The focus of future studies will test more machine learning methods including multiple runs with different combinations of training and test sets, to improve on results from this study and from [16].

Authors' contributions
MP collected the data, developed the methodology, performed the analysis, and wrote the manuscript. AM reviewed the manuscript. FP designed the study, developed the methodology, and reviewed the manuscript.

Competing interests
The authors declare that they have no competing interests.

References
1. Pirotti F, Guarnieri A, Vettore A. Collaborative Web-GIS design: a case study for road risk analysis and monitoring. Trans GIS. 2011;15:213–26.
2. Van Asselen S, Verburg PH. Land cover change or land-use intensification: Simulating land system change with a global-scale land change model. Glob Chang Biol. 2013;19:3648–67.
3. Berni JAJ, Zarco-Tejada PJ, Suárez L, Fereres E, Suarez L, Fereres E. Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. IEEE Trans Geosci Remote Sens. 2009;47(3):722–38.
4. Herwitz SR, Johnson LF, Dunagan SE, Higgins RG, Sullivan D V, Zheng J, et al. Imaging from an unmanned aerial vehicle: Agricultural surveillance and decision support. Comput Electron Agric. 2004;44(1):49–61.
5. Hunt ER, Dean Hively W, Fujikawa SJ, Linden DS, Daughtry CST, McCarty GW. Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring. Remote Sens. 2010;2(1):290–305.
6. Lelong CCD, Burger P, Jubelin G, Roux B, Labbé S, Baret F. Assessment of unmanned aerial vehicles imagery for quantitative monitoring of wheat crop in small plots. Sensors. 2008;8(5):3557–85.
7. Remondino F, Barazzetti L, Nex F, Scaioni M, Sarazzi D. UAV photgrammetry for mapping and 3D modeling – current status and future perspectives. In: Int Arch Photogramm Remote Sens Spat Inf Sci. 2011;38:14–16.
8. Papageorgiou EI, Markinos AT, Gemtos TA. Fuzzy cognitive map based approach for predicting yield in cotton crop production as a basis for decision support system in precision agriculture application. Appl Soft Comput. 2011;11(4):3643–57. https://doi.org/10.1016/j.asoc.2011.01.036.
9. Zheng YJ, Song Q, Chen SY. Multiobjective fireworks optimization for variable-rate fertilization in oil crop production. Appl Soft Comput. 2013;13(11):4253–63. http://dx.doi.org/10.1016/j.asoc.2013.07.004.
10. Hung C, Xu Z, Sukkarieh S. Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a UAV. Remote Sens. 2014;6(12):12037–54.
11. Guo X, Denman S, Fookes C, Mejias L, Sridharan S. Automatic UAV forced landing site detection using machine learning. In: 2014 Int Conf Digit Image Comput Tech Appl (DICTA), Wollongong, NSW, Australia, 25-27 November 2014. 2014. doi:10.1109/DICTA.2014.7008097.
12. Anthony D, Basha E, Ostdiek J, Ore J, Detweiler C. Surface Classification for Sensor Deployment from UAV Landings. In: 2015 IEEE Int Conf Robot Autom (ICRA), Seattle, WA, USA, 26-30 May 2015. 2015. doi:10.1109/ICRA.2015.7139678.
13. Cracknell MJ, Reading AM. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Comput Geosci. 2014;63:22–33.
14. Foody GM, Mathur A. A relative evaluation of multiclass image classification by support vector machines. IEEE Trans Geosci Remote Sens. 2004;42(6):1335–43.
15. Pal M. Random forest classifier for remote sensing classification. Int J Remote Sens. 2005;26(1):217–22.
16. Pirotti F, Sunar F, Piragnolo M. Benchmark of Machine Learning Methods for Classification of a Sentinel-2 Image. In: Int Arch Photogramm Remote Sens Spat Inf Sci XLI-B7:335–34. 2016. doi:10.5194/isprs-archives-XLI-B7-335-2016.
17. Song X. Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image. Int J Remote Sens. 2012;33(10):3301–20. doi:10.1080/01431161.2011.568531.
18. Waske B, Benediktsson JA, Árnason K, Sveinsson JR. Mapping of hyperspectral AVIRIS data using machine-learning algorithms. Can J Remote Sens. 2009;35(Suppl. 1):S106–S116.
19. Cortez P. rminer: Data Mining Classification and Regression Methods. 2016. Available from: https://cran.r-project.org/package=rminer. Accessed 27 June 2017.
20. Cortez P. Package "rminer" 2016. In: rminer: Data Mining Classification and Regression Methods. 2016. https://cran.r-project.org/web/packages/rminer/rminer.pdf. Accessed 27 June 2017.
21. Cortez P. Data Mining with Neural Networks and Support Vector Machines Using the R/rminer Tool. In: Perner P. (eds) Advances in Data Mining. Applications and Theoretical Aspects. ICDM 2010. Lecture Notes in Computer Science, vol 6171. Springer, Berlin, Heidelberg. 2010. https://link.springer.com/chapter/10.1007/978-3-642-14400-4_44. Accessed 27 June 2017.
22. Cortez P. A tutorial on using the rminer R package for data mining tasks, Teaching Report. Department of Information Systems, ALGORITMI Research Centre, Engineering School, University of Minho, Guimar˜aes, Portugal. 2015.
23. Cortes C, Vapnik V. Support-Vector Networks. Mach Learn. 1995;20:273–97.
24. Shi T, Horvath S. Unsupervised Learning With Random Forest Predictors. J Comput Graph Stat. 2006;15:118–38. http://www.tandfonline.com/doi/abs/10.1198/106186006X94072. Accessed 27 June 2017.
25. Breiman L. Random forests. Machine Learning. 2001;45(1):5–32. doi:10.1023/A:1010933404324.
26. Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2/3(December). 2002;18–22.
27. Gader P, Zare A, Close R, Aitken J, Tuell G. MUUFL Gulfport Hyperspectral and LiDAR Airborne Data Set. Univ. Florida, Gainesville, FL, USA, Tech. Rep. REP-2013-570. 2013. https://github.com/GatorSense/MUUFLGulfport. Accessed 27 June 2017.