LAND USE AND LAND COVER CLASSIFICATION USING HYPERSPECTRAL IMAGERY: EVALUATING THE PERFORMANCE OF SPECTRAL ANGLE MAPPER, SUPPORT VECTOR MACHINE AND RANDOM FOREST

L. E. Christovam[1*], G. G. Pessoa[1], M. H. Shimabukuro[1,2], M. L. B. T. Galo[1,3]

[1]Graduate Program in Cartographic Sciences, São Paulo State University, Presidente Prudente, Brazil
[2] Dept. of Mathematics and Computer Sciences, São Paulo State University, Presidente Prudente, Brazil
[3]Dept. of Cartography, São Paulo State University, Presidente Prudente, Brazil
{luiz.christovam, guilherme.g.pessoa, milton.h.shimabukuro, trindade.galo}@unesp.br

KEY WORDS: LULC, HyRANK dataset, Hyperspectral Classification, SAM, SVM, RF

ABSTRACT:

Land Use and Land Cover (LULC) information is an important data source for modeling environmental variables, so it is essential to develop high quality LULC maps. The hundreds of continuous spectral bands gathered with hyperspectral sensors provide high spectral detail and consequently confirm hyperspectral remote sensing as an appropriate option for many LULC applications. Despite increased spectral detail, issues like high dimensionality, huge volume of data and redundant information, mean that hyperspectral image classification is a complex task. It is therefore essential to develop classification approaches that deals with these issues. Since classification results are directly dependent on the dataset used, it is fundamental to compare and validate the classification approaches in public datasets. With this in mind, aiming to provide a baseline, four classification models in the relatively new hyperspectral HyRANK dataset were evaluated. The classification models were defined with three well-known classification algorithms: Spectral Angle Mapper (SAM), Support Vector Machine (SVM) and Random Forest (RF). A classification model with SAM and another with RF were defined with the 176 surface reflectance bands. A dimensionality reduction with principal component analysis was carried out and a classification model with SVM and another with RF were defined using 14 principal components as features. The results show that SVM and RF algorithms outperformed by far the SAM in terms of accuracy, and that the RF is slightly better than the SVM in this respect. It is also possible to see from the results that the use of principal components as features provided an improvement in the accuracy of the RF and an improvement of 28% in the time spent fitting the classification model.

1. INTRODUCTION

The pace of changes in Land Use and Land Cover (LULC) due to human activity are unprecedented. Activities pursuing economic objectives, such as wood exploration, cattle ranching and agriculture, drive LULC changes mainly through deforestation in tropical countries. LULC changes increase the probability of erosion and flooding occurrence and lead to problems such as loss of biodiversity and increase in greenhouse gases emissions (Mas, 1999). LULC information therefore has a key role in environmental and climate change studies (Henderson-Sellers and Pitman 1992), and the requirement of reliable and high quality LULC maps is a global concern, since these provide baseline information for planning and evaluating natural resource management, modeling environmental variables and developing sustainable practices (Adam et al., 2014; Gómez et al., 2016).

Since the 1970s, remote sensing imagery has provided an uninterrupted and reliable set of information enabling the mapping and monitoring of the Earth's surface (Petitjean et al., 2012; Maus, et al., 2016). The synoptic and multitemporal characteristics, as well as the large coverage area, provide capabilities that make satellite imagery the most suitable approach for mapping large areas, taking account of time and expense (Kavzoglu and Colkesen, 2009, Puletti et al., 2016). While remote sensing satellites have the capability of capturing landscape imagery, extracting LULC information from these data requires effective image processing techniques.

Although satellite images have been used for decades for LULC mapping, and the remote sensing community has been seeking improvements in image classification techniques throughout that time, there is still room for improvement in LULC mapping through digital image processing. The construction of LULC products through image classification techniques is a challenging task, due to several variables related to the process (Shao and Wu, 2008, Manandhar et al., 2009, Khatami et al., 2016). Among the several elements that influence the accuracy of remote sensing image classification, defining the image and the classification algorithm are particular issues.

A few decades ago, hyperspectral remote sensing arose offering high potential to improve LULC mapping (Biocudas-Dias et al., 2013). The possibility of acquiring hundreds of continuous spectral bands with narrow bandwidths, providing more spectral detail than coarse bandwidths acquired with multispectral sensors, provided new opportunities for LULC applications (Chutia et al., 2016; Ghamisi et al., 2017). However, higher spectral detail comes with high dimensionality and a huge volume of data, which leads to an issue known as the Hughes phenomenon. This explains that increasing the number of predictor features potentially adds information to separate the classes, however the complexity is also increased and the number of samples in short dataset may not be enough to characterize this complexity, therefore adding more features rather than increasing the classification accuracy may actually decrease the accuracy (Maxwell et al., 2018).

---

[*] Corresponding author

Hyperspectral image classification is also a challenging task due to the limited number of available training samples, the redundant information presented in the features, the uncertainties related to atmospheric or topographic effects and the influence of spatial resolution (Ghamisi et al., 2017). Choosing an algorithm that deals with these issues is therefore essential to get high accuracy LULC products through hyperspectral image classification. Nowadays, machine learning algorithms like Support Vector Machine (SVM), Random Forest (RF) and Neural Nets (NN) have been widely used by the remote sensing community for hyperspectral image classification (Ghamisi et al., 2017), although the results provided by each are dependent on the dataset used, therefore studies that evaluate and compare them in the same dataset are of particular interest.

For this reason, and due to the lack availability of public hyperspectral datasets of large size and with several land cover classes, the *International Society for Photogrammetry and Remote Sensing* (ISPRS), commission III, working group III/4, hyperspectral image processing, developed the HyRANK Hyperspectral Satellite Dataset (Karantzalos, et al., 2018), aiming to provide a dataset where researchers will be able to validate and compare new LULC classification approaches.

Based on that, the accuracy of four classification models defined with three well-known algorithms carrying out the classification on the HyRANK dataset were evaluated and compared in this paper. The goal was to provide a baseline for future research in LULC classification with the HyRANK dataset. The algorithms adopted were the traditional Spectral Angle Mapper (SAM), which in the recent past has been widely used in the classification of hyperspectral data, and two state-of-the-art machine learning algorithms, SVM and RF. This paper is organized as follows: an overview of the classification algorithms is presented in section 2, in section 3 the dataset, the classification process and the metrics used in the evaluation process are explained, in section 4 the results and discussion are produced. Finally, the conclusions are given in section 5.

2. CLASSIFICATION ALGORITHMS

A short overview of the classification algorithms SAM, SVM and RF is presented in this section with the aim of providing a background to these classifiers to help in understanding the sections which follow. More details about SAM, SVM and RF can be found, respectively, in Kruse et al. (1993), Cortes and Vapnik (1995), and Breiman (2001).

SAM is a non-parametric supervised algorithm proposed in Kruse et al. (1993). The algorithm measures the similarity between two spectra, where one corresponds to the pixel to be labeled and the other is a reference spectrum for the LULC class. The spectra are treated as vectors, and their dimension is defined by the number of image bands. The similarity measurement is the angle between the two spectra from the origin, and it can be computed with Equation 1, where $\vec{t}$ is the pixel spectrum and $\vec{r}$ the spectrum for the LULC class.

$$\theta = \cos^{-1}\left(\frac{\vec{t}\cdot\vec{r}}{\|\vec{t}\|\cdot\|\vec{r}\|}\right) \qquad (1)$$

SAM is less sensitive to illumination variations than other similarity measurements (i.e. Euclidean distance), since the changes in illumination have a direct effect on the scale and an indirect effect on the orientation, considering the line that connects the pixel and the origin of the multidimensional feature

space. In this case, even with changes in the pixel digital number, it will be aligned with its class (Kruse et al., 1993). However, the need for pure reference spectra is the major drawback of this classifier.

Machine learning classifiers have become a major focus for the remote sensing community since such algorithms deal with high-dimensional feature space and are able to model complex class signatures (Maxwell et al., 2018). Nowadays the non-parametric machine learning algorithms SVM, RF and NN are state-of-the-art for remote sensing image classification (Khatami et al., 2016), producing higher levels of accuracy than parametric algorithms like Maximum Likelihood (Yu et al., 2014).

SVM finds the optimal hyperplane that separates the classes in a multi-dimensional feature space. The best decision boundary is that which minimizes the errors and maximizes the distance between the training samples (Cortes and Vapnik, 1995). SVM is especially useful for small training datasets since it relies only on observations located on the decision boundaries (support vectors) (Mountrakis et al., 2011). This advantage makes SVM more relevant for remote sensing applications in particular, due to the problem of getting training samples which normally require field work and consequently high cost (Tuia et al., 2011).

Despite the advantages, SVM has a drawback in the high number of parameters to be tuned, a high computational cost and the need to choose a kernel function (Mountrakis et al., 2011). The kernel function is used to transform the n-dimensional feature space into a larger dimension space where the classes are linearly separable. The known kernel options are linear, polynomial, sigmoid, and the radial-base function (RBF) (Kavzoglu and Colkesen, 2009). Mountrakis et al. (2011) highlights that RBF kernel is the one most suitable for remote sensing data classification.

RF algorithm is an ensemble of decision trees. It has been used in several remote sensing classification works because of its simplicity and good accuracy results (Belgiu and Drăguţ, 2016).

Decision tree is a recursive split approach of the input data (Pal and Mather, 2003). The splits are performed starting from a root node (first level of the tree) up to the leaf nodes, decreasing the entropy at each split. The leaves are the last level of the tree and it is where the entropy is at its lowest possible. The intention is to have only samples from the same class in the leaves (Ho, 1995). There are several split nodes in the path that goes from the root node to the leaf node. These contain decision rules based on the available features and a threshold applied to the features chosen.

Despite the decision trees being extremely fast and simple, they are very sensible to noise and frequently overfit the training samples. Because of that, a decision tree can be classified as a weak learner. To overcome these drawbacks an ensemble of decision trees is combined in an RF (strong learner). In the RF algorithm the trees of the forest must be uncorrelated, each tree being unique, hence the random subspace (feature bagging) and bootstrap aggregating (bagging) techniques are applied.

Bootstrap aggregating, presented by Breiman (2001), consists of the random selection, with replacement, of a subset of samples from the training dataset. The random subspace, proposed by Ho (1995), consist of a randomly selected subset of features from all the input features at each node, and, from the new subset of features chosen, considering the one which splits the node that produces the smaller entropy at the next level.

As stated before, both techniques minimize the model variance without increasing the trend. So, while a single decision tree is sensitive to noise, the average forecast for an ensemble of trees is not sensitive, so long as the trees are uncorrelated (Friedman et al., 2001). After growing the forest, each tree casts a vote for a class and the label is defined by the majority vote. The main advantages of the RF algorithm are in dealing well with noise, having fewer parameters to be tuned, and the low computational cost.

3. MATERIAL AND METHODS

3.1 Dataset

As stated before, in this work the HyRANK dataset, developed by a scientific initiative of the ISPRS, WG III/4, was used. The main goal with the HyRANK dataset is to provide a dataset with hyperspectral images along correspondent ground truths, to enable the scientific community to validate new classification approaches against the state-of-the-art methods.

HyRANK is composed of five hyperspectral images gathered with the Hyperion sensor Earth Observing-1. After a pre-processing step, these images were provided with 176 surface reflectance bands. The ground truths from two (Loukia and Dioni) of five images were provided. 14 LULC classes were annotated in the ground truth following the CORINE Land Cover principles. These classes are: Dense Urban Fabric, Mineral Extraction Sites, Non-Irrigated Arable Land, Fruit Trees, Olive Groves, Broad-leaved Forest, Coniferous Forest, Mixed Forest, Dense Sclerophyllous Vegetation, Sparse Sclerophyllous Vegetation, Sparsely Vegetated Areas, Rocks and Sand, Water and Coastal Water. The Loukia and Dioni images, as well as the respective ground truths, are depicted in Figure 1.

As well as being a dataset, HyRANK is intended to be an evaluation online platform in which it will be possible to upload the LULC results for the images Erato, Nefeli and Kiriki. The overall accuracy of the model will be estimated in this platform and will be made available to the community in order to compare results from different classification approaches. Further details regarding the dataset and the accuracy assessment platform can be found in Karantzalos et al. (2018).

At the time of writing, the images are already available, but the online assessment platform is not yet. In order to obtain, train and evaluate classification models, the ground truths from Loukia and Dioni images were therefore split in two sets. 85% of the ground truth were assigned for training the classification algorithms, and the remaining 15% were applied to classification model evaluation.

3.2 LULC classifications description

All steps to implement the LULC classification and evaluation were performed using available libraries in python. The SAM algorithm is available in the *spectral* package and the machine learning algorithms are available in the *scikit-learn* package (Pedregosa et al., 2011).

To perform the SAM's classification we used the 176 surface reflectance bands as features, so each class and pixel spectrum has a 176-dimension. The 14 LULC classes had their reference spectra defined based on the training set, all reference spectra were assembled through the mean of all pixels' spectra belonging to the correspondent LULC class. In the classification process, each pixel spectrum was compared to the 14 reference spectra, so that 14 similarity measurements were produced for each pixel. The label of each pixel was defined as the same as the most similar (lower angle) reference spectrum.

As stated in Section 2, the SVM algorithm has high computational cost, which makes it not feasible to use 176 surface reflectance bands as input features. Another issue related to the feature dimension is the Hughes phenomenon, which can lead to a decrease in the accuracy of the classification models as stated in Section 1. A dimensionality reduction was therefore carried out in the dataset using the principal components analysis (PCA). More details about PCA can be found in Johnson and Wichern (2002). The PCA parameters were fitted over the training samples, but, before that, the training samples were balanced, since unbalanced classes could influence the eigenvectors of the PCA. Each of the 14 LULC classes was balanced through replication. In the sequence, the PCA transformation was performed over the dataset, verifying that 99.9% of the information was presented in 14 principal components, these components were selected as significant input features.

For the SVM classification it was decided to use the RBF kernel, since it is the most suitable option for remote sensing classification. Hence, knowing that SVM with this kernel option

(1) Dense Urban Fabric
(2) Mineral Extraction Sites
(3) Non Irrigated Arable Land
(4) Fruit Trees
(5) Olive Groves
(6) Broad_leaved Forest
(7) Coniferous Forest
(8) Mixed Forest
(9) Dense Sclerophyllous Vegetation
(10) Sparce Sclerophyllous Vegetation
(11) Sparcely Vegetated Areas
(12) Rocks and Sand
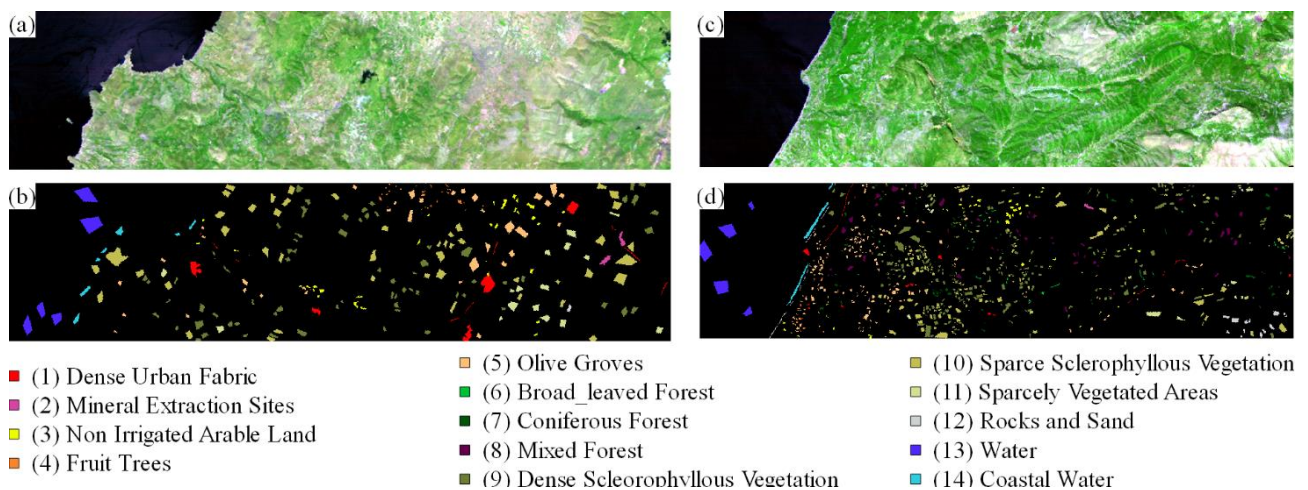(13) Water
(14) Coastal Water

Figure 1. (a) Dioni image composition B25G90R160 (b) Dioni ground truth (c) Loukia image composition B25G90R160 (d) Loukia ground truth

assumes all features are centered around 0 and have variance in the same order, the 14 principal components were standardized by removing the mean and scaling to unit variance before the training step. In the sequence, the SVM training was performed using the 14 significant principal components as input features. The classification model having been defined, the classification was carried out and the LULC maps produced for Loukia and Dioni images. The *SVC* class from the *scikit-learn* package was used to perform the multiclass classification with SVM. As stated, the RBF kernel function was used, the *class_weight* was defined as *balanced* and other parameters were set as default.

Since it was desirable to compare the RF algorithm with the other classifiers and verify whether the Hughes phenomenon would disturb the accuracy of the models, two classification models were defined with the RF algorithm. In the first classification model the 176 surface reflectance bands were used and in the second the input features were 14 significant principal components. The classification was performed with the 176 features since it was known that the RF has low computational cost, so using all available features would not be a time problem, as it is with SVM. The two classification models defined, the classifications were carried out and the LULC maps produced for Loukia and Dioni images. For this work the *RandomForestClassifier* class from the *scikit-learn* package was used. The number of trees was set equal to 200 and the depth equal to 20. These values were defined after a tuning process through which it was realized that the overall accuracy would not improve significantly, thereby increasing these parameter values.

3.3 Assessment of the accuracy of classification models

Metrics computed from the confusion matrix were used for the evaluation of the classification models. 15% of samples from the Dioni and Loukia ground truths were applied in the construction of the confusion matrix. These samples were randomly selected before definition of the classification models, so these samples were not used to fit the classification models. The metrics applied for the classifications accuracy assessment were: Overall Accuracy (OA), Producer's Accuracy (PA), User's Accuracy (UA), Kappa coefficient (K) and F1-score. The OA is a metric for general evaluation, which is computed through the sum of correctly classified pixels divided by the total of pixels in the validation set. The PA, UA and F1-score metrics were used to measure the quality of the classification for each LULC class. The PA indicates the omission of pixels for a class, the UA expresses the commission of pixels and the F1-score is the harmonic mean between the PA and UA, so it represents how well the model classifies each class. The K coefficient as the OA is a measure of general evaluation, although, it can be considered a more representative metric since wrongly classified samples are considered in its computation. Further details regarding the metrics used in the classification accuracy assessment can be found in Congalton (1991). Aiming to compare the classifiers, the classification model fit time was also measured for each of the 4 classifications performed. All experiments were carried out on a machine with an Intel Core i7-5500, 2.4 GHz clock and 8Gb RAM.

4. RESULTS AND DISCUSSION

Figure 2 shows the LULC maps for Dioni (left) and Loukia (right) images produced through 4 different classification models as described in Section 3.2. The metrics used for the accuracy assessment of the classification models, as well as the time to fit them, are shown in Table 1.

The most noticeable issue in the LULC maps appeared through a visual analysis of the Dioni LULC map built with the SAM algorithm, Figure 2 (a). The Dense Urban Fabric class was
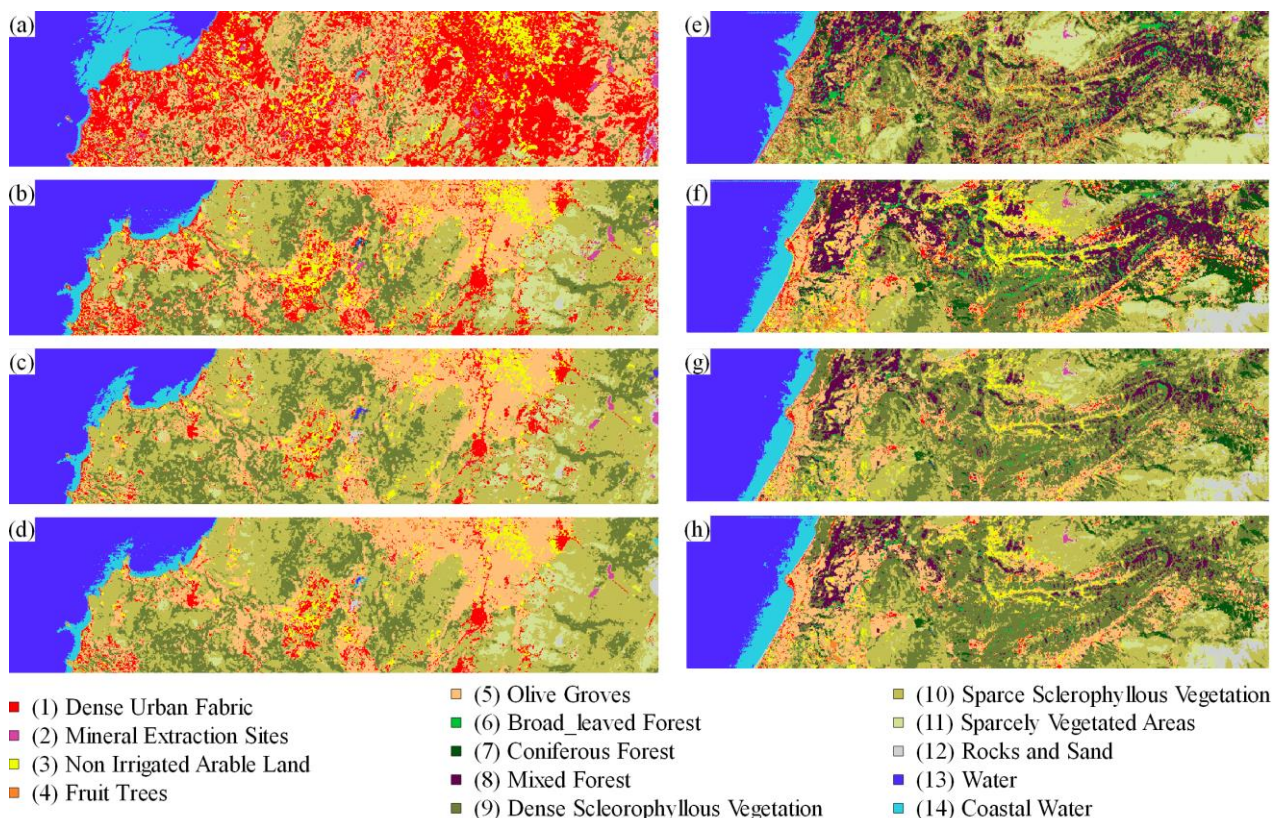

- (1) Dense Urban Fabric
- (2) Mineral Extraction Sites
- (3) Non Irrigated Arable Land
- (4) Fruit Trees
- (5) Olive Groves
- (6) Broad_leaved Forest
- (7) Coniferous Forest
- (8) Mixed Forest
- (9) Dense Scleorophyllous Vegetation
- (10) Sparce Sclerophyllous Vegetation
- (11) Sparcely Vegetated Areas
- (12) Rocks and Sand
- (13) Water
- (14) Coastal Water

Figure 2. LULC classification for Dioni image through (a) SAM (b) SVM (c) RF_bands (d) RF_pca – LULC classification for Loukia image through (e) SAM (f) SVM (g) RF_bands (h) RF_pc

	SAM			SVM			RF_bands			RF_pca		
Overall Accuracy	0.48			0.88			0.91			0.92		
K coefficient	0.41			0.86			0.89			0.91		
Fit time	61.38 s			123.38 s			43.77 s			31.2 s		
	PA	**UA**	**F₁**	**PA**	**UA**	**F₁**	**PA**	**UA**	**F₁**	**PA**	**UA**	**F₁**
Dense Urban Fabric	0.15	0.58	0.23	0.85	0.90	0.87	0.93	0.78	0.85	0.94	0.83	0.88
Mineral Extraction Sites	0.29	0.83	0.43	1.00	0.97	0.98	0.93	0.78	0.85	1.00	0.88	0.93
Non-Irrigated Arable Land	0.66	0.66	0.66	0.81	0.93	0.87	0.92	0.89	0.91	0.94	0.91	0.92
Fruit Trees	0.11	0.49	0.18	0.61	0.89	0.73	0.91	0.54	0.68	0.96	0.62	0.75
Olive Groves	0.19	0.23	0.21	0.95	0.88	0.91	0.91	0.90	0.91	0.91	0.96	0.94
Broad-leaved Forest	0.33	0.77	0.46	0.57	0.88	0.69	0.90	0.68	0.77	0.92	0.60	0.73
Coniferous Forest	0.46	0.77	0.58	0.78	0.90	0.84	0.96	0.86	0.90	0.96	0.83	0.89
Mixed Forest	0.30	0.68	0.42	0.47	0.87	0.61	0.74	0.65	0.69	0.79	0.68	0.73
Dense Sclerophyllous Veg.	0.77	0.48	0.59	0.92	0.81	0.86	0.90	0.91	0.90	0.90	0.93	0.92
Sparce Sclerophyllous Veg.	0.79	0.36	0.50	0.92	0.89	0.91	0.89	0.95	0.92	0.92	0.95	0.94
Sparcely Vegetated Areas	0.29	0.25	0.27	0.88	0.93	0.90	0.91	0.90	0.90	0.96	0.92	0.94
Rocks and Sand	0.42	0.35	0.38	0.99	0.96	0.97	0.96	0.94	0.95	1.00	0.96	0.98
Water	0.97	0.89	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Coastal Water	0.69	0.89	0.78	1.00	1.00	1.00	1.00	0.98	0.99	1.00	0.99	1.00

Table 1 – Evaluation metrics used for evaluation of the classification approaches, Overall Accuracy (OA), Kappa index (K), Model fit time, Producer's Accuracy (PA), User's Accuracy (UA) e F₁-score

overestimated, since the area estimated for this class is larger than its real area, as can be seen in Figure 1 (a). However, this is not an unexpected result considering that this class embraces different land cover types (i.e. vegetation and bare soil) which can diminish its classification accuracy, since SAM requires pure reference spectra to work well. Another noticeable issue in this LULC map is the confusion of the Water class with Coastal Water, which can be explained by the similar spectral signatures of these classes.

With regard to the metrics generated for accuracy assessment of the classification models, it can be seen that the two models defined with the RF algorithm presented the best performances, with a K coefficient equal to 0.91 and 0.89 for the principal components and the 176 surface reflectance bands as input features respectively. Despite the almost insignificant improvement, a slightly better performance was achieved using principal components as features. This can be related to the Hughes phenomenon, since the model complexity is higher for the 176 features than for the 14 features. The RF classification model performances were followed by the model fitted with SVM which had a K coefficient equal to 0.86 and SAM's classification model with K coefficient equivalent to 0.41. So, as expected, both machine learning algorithms outperformed by far the classification with SAM. Again, this can be related to SAM's need for pure reference spectra, which is almost impossible to get for moderate spatial resolution imagery like the Hyperion sensor which has spatial resolution of 30 meters

In general, UA and PA for the machine learning algorithms were around or more than 0.80 for almost every class, the exceptions being the Fruit Trees, Broad Leaved Forest and Mixed Forest classes. The Mixed Forest class presented a PA of 0.47 for SVM, PA and UA of 0.74 and 0.65 respectively for RF_bands; and UA of 0.68 for RF_pca. The Fruit Trees class presented a PA of 0.61 for SVM, with a UA of 0.54 and 0.62 for RF_bands and RF_pca respectively. The Broad Leaved Forest class presented a PA of 0.57 for SVM and a UA of 0.68 and 0.60 for RF_bands and RF_pca respectively. For the SAM, the UA and PA presented lower values achieving averages of 0.58 and 0.45, respectively. The exception, in this case, was the Water class which reached a PA of 0.97 and UA of 0.89, these values are close to the UA and PA achieved with the machine learning algorithms, which can be explained by two facts. Firstly, the spectral signature for the

Water class is very similar to the Coastal Water class, but extremely different to the other classes. Secondly, it could be linked to the higher likelihood of getting a pure reference spectrum for the Water, which is essential to SAM.

Regarding the F1-score for SAM's classification model, the Dense Urban Fabric, Fruit Trees, Olive Groves and Sparsely Vegetated Areas classes were under 0.3 showing that the boundary limits in the features space were poorly defined for these classes. In general, the RF_pca classification model performed better than RF_bands and SVM classification models, since the OA for the former was slightly higher than for the latter. Although, looking for the F1-score, it can be seen that the Broad Leaved Forest and Coniferous Forest classes had their boundary limits better defined in the feature space by the RF_bands classification model, while the Mineral Extraction Sites class was better classified with SVM, since its F1-score was higher.

Considering the computational time, the RF algorithm again outperformed the other algorithms, the RF_pca was the fastest taking just 31.2s to fit. This was followed by the RF_bands with 43.7s. The SAM and SVM were, respectively, almost twice and four times slower than RF_pca. However, taking into account just the time to fit the classification models, if the time to perform the PCA is considered, the RF_pca and SVM would perform slower than the others.

5. CONCLUSION

In this work, the evaluation of four classification models defined with three well-known classification algorithms, SAM, SVM and RF, was assessed for the LULC classification using the public HyRANK dataset. The results show that the SVM and RF algorithms, outperformed by far the SAM in terms of accuracy of the classification models and that the RF is slightly better than the SVM in the same respect. From the results it can also be seen that the of use principal components as features provided a slight improvement in the accuracy of the RF and an improvement of 28% in the time spent to fit the classification model.

Although the results obtained for the machine learning algorithms, RF and SVM, are considered excellent, with a K coefficient higher than 0.8, there is still room for improvement. One possibility is to perform the selection of bands that best

contribute to the delimitation of the classes in the feature space so that the importance of each attribute generated by the RF algorithm can be used. Another possibility for future work is to use deep NN in this dataset, since it can improve the results because of its capacity for engineering new features. The results presented in this paper can be used as a baseline for future research in hyperspectral image classification on HyRANK dataset.

ACKWNOLEGDMENTS

REFERENCES

Adam, E., Mutanga, O., Odindi, J. and Abdel-Rahman, E.M., 2014. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, *35*(10), pp.3440-3458.

Belgiu, M. and Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, pp.24-31.

Bioucas-Dias, J.M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N. and Chanussot, J., 2013. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and remote sensing magazine*, *1*(2), pp.6-36.

Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp.5-32.

Chutia, D., Bhattacharyya, D.K., Sarma, K.K., Kalita, R. and Sudhakar, S., 2016. Hyperspectral remote sensing classifications: a perspective survey. *Transactions in GIS*, *20*(4), pp.463-490.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment*, *37*(1), pp.35-46.

Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine learning*, *20*(3), pp.273-297.

Friedman, J., Hastie, T. and Tibshirani, R., 2001. *The elements of statistical learning* (Vol. 1, No. 10). New York: Springer series in statistics.

Ghamisi, P., Plaza, J., Chen, Y., Li, J. and Plaza, A.J., 2017. Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geoscience and Remote Sensing Magazine*, *5*(1), pp.8-32.

Gómez, C., White, J.C. and Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, pp.55-72.

Henderson-Sellers, A. and Pitman, A.J., 1992. Land-surface schemes for future climate models: Specification, aggregation, and heterogeneity. *Journal of Geophysical Research: Atmospheres*, *97*(D3), pp.2687-2696.

Ho, T.K., 1995, August. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278-282). IEEE.

Johnson, R.A. and Wichern, D.W., 2002. *Applied multivariate statistical analysis* (Vol. 5, No. 8). Upper Saddle River, NJ: Prentice hall.

Karantzalos, Konstantinos, Karakizi, Christina, Kandylakis, Zacharias, & Antoniou, Georgia. (2018). HyRANK Hyperspectral Satellite Dataset I (Version v001) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.1222202

Kavzoglu, T. and Colkesen, I., 2009. A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, *11*(5), pp.352-359.

Khatami, R., Mountrakis, G. and Stehman, S.V., 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment*, *177*, pp.89-100.

Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J. and Goetz, A.F.H., 1993. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, *44*(2-3), pp.145-163.

Manandhar, R., Odeh, I. and Ancev, T., 2009. Improving the accuracy of land use and land cover classification of Landsat data using post-classification enhancement. *Remote Sensing*, *1*(3), pp.330-344.

Mas, J.F., 1999. Monitoring land-cover changes: a comparison of change detection techniques. *International journal of remote sensing*, *20*(1), pp.139-152.

Maus, V., Câmara, G., Cartaxo, R., Sanchez, A., Ramos, F.M. and de Queiroz, G.R., 2016. A time-weighted dynamic time warping method for land-use and land-cover mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(8), pp.3729-3739.

Maxwell, A.E., Warner, T.A. and Fang, F., 2018. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, *39*(9), pp.2784-2817.

Mountrakis, G., Im, J. and Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(3), pp.247-259.

Pal, M. and Mather, P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote sensing of environment*, *86*(4), pp.554-565.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, *12*(Oct), pp.2825-2830.

Petitjean, F., Kurtz, C., Passat, N. and Gançarski, P., 2012. Spatio-temporal reasoning for the classification of satellite image time series. *Pattern Recognition Letters*, *33*(13), pp.1805-1815.

Puletti, N., Camarretta, N. and Corona, P., 2016. Evaluating EO1-Hyperion capability for mapping conifer and broadleaved forests. *European Journal of Remote Sensing*, *49*(1), pp.157-169.

Shao, G. and Wu, J., 2008. On the accuracy of landscape pattern analysis using remote sensing data. *Landscape Ecology*, *23*(5), pp.505-511.

Tuia, D., Volpi, M., Copa, L., Kanevski, M. and Munoz-Mari, J., 2011. A survey of active learning algorithms for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Signal Processing*, *5*(3), pp.606-617.

Yu, L., Liang, L., Wang, J., Zhao, Y., Cheng, Q., Hu, L., Liu, S., Yu, L., Wang, X., Zhu, P. and Li, X., 2014. Meta-discoveries from a synthesis of satellite-based land-cover mapping research. *International Journal of Remote Sensing*, *35*(13), pp.4573-4588.