NEW APPROACH OF SAMPLE GENERATION AND CLASSIFICATION FOR WILDFIRE FUEL MAPPING ON HYPERSPECTRAL (PRISMA) IMAGE

Riyaaz Uddien Shaik[1], Laneve Giovanni[2], Lorenzo Fusilli[1]

[1]Dept. of Astronautical, Electrical and Energy Engineering (DIAEE)*, Sapienza University of Rome.
[2]School of Aerospace Engineering, Sapienza University of Rome, Italy.

ABSTRACT

Hyperspectral images have its applications in various fields. Here, hyperspectral image from PRISMA which is a fundamental satellite of Italian Space Agency is being used for discriminating the wildfire fuel types on Sardinian Island of Italy. PRISMA is an on-demand mission and the available data in the archive are limited. There is no literature available on land use/vegetation classification using PRISMA data. In this paper, a new approach for generating samples to form a dataset for classifying the wildfire fuels and for classifying mixed pixels using iso-bioclimatic conditions are proposed. The classified map created using the dataset and using the iso-bioclimatic conditions is been validated. From the accuracy assessment, SVM classifier showed an overall accuracy of 86% and kappa coefficient of 0.79. Then, the classified map is converted into fuel map. This study suggests that the proposed approach can be used to generate samples for land use/vegetation classification and to assign vegetation types to mixed pixels depending upon the iso-bioclimatic conditions.

Index Terms—*PRISMA, Hyperspectral, Guided Filtering, SAMSID Mixed Measure.*

1. INTRODUCTION

Remote sensors from satellites, drones etc., in this new generation are producing hyperspectral images from the visible to shortwave infrared parts of the electromagnetic spectrum in sample hundreds of contiguous narrow bands. PRISMA (from PRecursore IperSpettrale della Missione Applicativa) launched by the Italian Space Agency (ASI) in 2019 has taken us in to a new era of hyperspectral imaging spectroscopy. In this work, hyperspectral imagery from PRISMA satellite is being used for mapping the vegetation fuel types. Access to the data has been granted to users in early June 2020 after the commissioning phase. A continuum of spectral bands with 400 to 2500 nm at a spatial resolution of 30 m can be captured using the imaging spectrometer. Totally, 173 bands in the shortwave infrared (SWIR) within 920-2500 nm and 66 bands are in the visible near-infrared (VNIR) portion of the spectrum (400-1010 nm). The widths and spectral sampling intervals are ≤12 nm. A panchromatic camera providing single band (400 – 700 nm) image at 5 m spatial resolution is also onboard the ASI's satellite. PRISMA is an on-demand mission and the available data in the archive are limited[1].

The available techniques/datasets are not suitable for all kinds of hyperspectral data, so development of new techniques/datasets for analysis of high dimensional data is necessary. There are various approaches developed to generate or to improve existing datasets such as GAN, cascade GAN etc., for hyperspectral image processing. In this work, we proposed a new methodology of sample generation for classifying fuel types using Support Vector Machine (SVM). SVM had significant advantages in accuracy, simplicity, and robustness[2]. It can produce better accuracy even with lesser amount of training data but has the limitation that it can be used only for binary classification. Possible solution is by training several SVM's simultaneously in all, all-against-one or one-against-one scheme [3]. Other important point to be considered in hyperspectral data classification is about classifying mixed pixels. In our case, iso-bioclimatic map is used to assign mixed pixels to their appropriate category.

In this work, wildfire fuels were detected using the hyperspectral imagery. Detection of fuel types helps in generating fuel map which could help fire managers, researchers, policy makers and systems in applications such as study of fire behaviors, fire potential, fire emissions, carbon budget, fuel management, fire effects and ecosystem modelling. In the framework of the EFFIS (European Forest Fire Information System) and FUELMAP projects, fuel map was generated categorizing 42 types in to 9 groups as Peat bogs, Grasslands, Shrublands, Transitional Shrubland/Forest, Coniferous forest, Broadleaved forest, Mixed forest, Marshes, riparian and coastal vegetation and Agro-Forestry areas. In this project, vegetation types from CORINE land cover were considered and categorized into fuel types referring to the EFFIS project of JRC [4][5].

2. METHODOLOGY

For classification of fuel types, a certain procedure was followed to prepare dataset by generating samples and then SVM classifier was applied. Collecting in-situ data was impossible due to COVID pandemic so by knowing single pixel, similar pixel wise vectors were formed using SAMSID-GF with k-means clustering methodology for dataset generation.

2.1. SAMSID

Spectral Angle Mapper (SAM) and Spectral Information Divergence (SID) are two popular techniques in hyperspectral data processing. SAM extracts geometric features between two spectra whereas SID views each pixel spectrum as a random variable and then measures the discrepancy of probabilistic behaviors between two spectra. So, SAM-SID mixed measure was implemented in finding the similar spectral signatures. SAM-SID can be represented as in eqn (1).

$$SAMSID = SID * \sin(SAM) \tag{1}$$

SAMSID is a product of the SID and the sine of the SAM between two spectral signatures [6].

2.2. Guided Image Filtering

Sharpening of edges was carried out on the image obtained from SAMSID method. For this process, a novel type of explicit image filter – guided filter is being implemented. The guided filter generates the output by considering the content of a guidance image which can be the input image itself or another different image. Like the bilateral filter, guided filter can also perform smoothing operation preserving edges which could help in finding out the threshold of that specific class. Guided image I, an input image p, and an output image q is a general linear translation – variant filtering process that can be represented as in eqn (2).

$$q_i = \Sigma W_{ij} I p_j \qquad (2)$$

where i and j are pixel indexes. The filter kernel W_{ij} is a function of the guidance image I and independent of p. This filter is linear with respect to p. In our case, guidance image has 3 bands from the hyperspectral cube selected using principal component analysis [7].

2.3. K-means clustering analysis

Deciding the threshold for Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), Spectral Angle Mapper (SAM), Spectral Information Divergence (SID) etc., by virtual inspection is a difficult task. So, k-means cluster analysis which is one of the machine learning tools could help in deciding the threshold by itself. K-means is a popular machine learning algorithm used for clustering data where the number of clusters (K) is either known, presumed, or indicated beforehand. As per the suggestion of Nguyen et al. [8] it would be better if the number of clusters K is greater than the number of actual or expected classes, so three clusters were formed since it is a binary classification [9]. The k-means algorithm was carried out using the Machine Learning Toolbox of MATLAB.

2.4. Support Vector Machine for HSI Classification

There are numerous supervised learning-based algorithms in artificial intelligence field which can be applied for classification. In the current framework, SVM (with Radial Basis Function, RBF) was applied because of its reputation in training data set to achieve high accuracy and outstanding generalization capability. This method works on statistical learning theory and structural risk minimization principle. The optimal separating hyperplane with the maximum margin between the classes will be found using the strategy of this classifier using the training samples located at the edge of the class distribution. The formula for the output of a linear SVM is shown in eqn (3):

$$u = w.x - b \qquad (3)$$

Where w is the normal vector to the hyperplane and x is the input vector. The separating hyperplane is the plane u=0. The nearest points lie on the planes u = ±1. The margin m is shown in eqn (4) as

$$m = \frac{1}{||w||_2} \qquad (4)$$

Since it is a non-linear data, kernels are required for training the model. So, in our case, RBF kernels were implemented. RBF(Gaussian) kernels are a family of kernels where smoothening of distance measure is done by radian function (exponential function) as shown in eqn (5). This kernel nonlinearly maps samples into a higher dimensional space so it, unlike the linear kernel, can handle the case with relation between class labels and attribute is nonlinear.

$$k\left(x_i, x_j\right) = \exp\left(-\sigma||x_i - x_j||^2\right), \sigma > 0 \qquad (5)$$

The adjustable parameter plays a major role in the performance of the kernel and should be carefully tuned. If overestimated, the exponential will behave almost linearly whereas if underestimated, the function will lack regularization and the decision boundary will be highly sensitive to noise in training data. Boundary will be highly sensitive to noise in training data so the complete behaviour of SVM depends on the choice of the width parameter [10]. To find the optimal values, initially hyperparameter optimization in classification learner app of MATLAB was used but slowly lead to overfitting so bayesian optimization was used just to optimize few parameters such as sigma, box constraint etc.

3. IMPLEMENTATION

The steps followed to implement the methodology is as explained below:

3.1. Study Area

This work is a part of S2IGI (an Integrated System for Forest Fire Management) project funded by Sardinia Island of South Italy which is having the characteristics of frequent fire occurrences. As a demonstration purpose, image of southern part of this Mediterranean Island was selected due to the availability of suitable data. It comprises complete Cagliari city and a part of Monte-Arcosu forest, one of the biggest holm oak forests of the Mediterranean region.

3.2. Data Handling

Here, we rely only on VNIR bands as the combination of VNIR and SWIR bands doesn't give better accuracies for coarser spatial resolution [11], [12]. Moreover, in the VNIR part of the spectrum, signal-to-noise ratio is higher and atmospheric effects from water vapor absorption are lower than SWIR. Level 2C products from the PRISMA archive with minimal cloud cover were considered. Conversion from uint16 Digital Number (DN) to reflectance values (no units) in the range of [0-1] were carried out using the formula as in eqn (6):

$$Reflectance = b + DN * \left(\frac{a-b}{65535}\right) \qquad (6)$$

Where a = L2ScaleVnriMax (obtained from metadata of image), b = L2ScaleVnirMin (obtained from metadata of image)

The atmospheric correction of the level 2C products is based upon inverting the radiative transfer model, i.e., minimizing a cost function representing the difference between the simulated spectrum and the measured one. The simulations are performed by the MODTRAN model and they are stored as a lookup-table (LUT) to speed up the inversion. Geo-referencing of the image was carried out by creating geographic look-up table (GLT) using the latitude and longitude values given in the metadata of the image. Further details about the mission and products are available in the PRISMA products specification document.

3.3. Removal of bad bands

For most hyperspectral remote sensing applications, removing bad bands is a required preprocessing step.

Currently, the commonly applied method is by visual inspection, which is very time-consuming, but it is easy to overlook some noisy bands. The 58 spectral bands in the 400–900 nm range, with a spectral bandwidth within ~9–12 nm, were extracted from L2C VNIR data to perform the study leaving 8 other bad bands.

3.4. Pixel Extraction

Pixels extraction of the HSI (Hyperspectral Image) is one of the imperative pre-processing mechanisms. It assists to handle the data and to implement the machine learning algorithms such as clustering, classification, etc. The individual elements in the HSI are pixels which are vectors. Pixels that correspond to fuel types by comparison with CLC 2018 and other data available from Copernicus are selected to form as vectors.

3.5. Dataset preparation

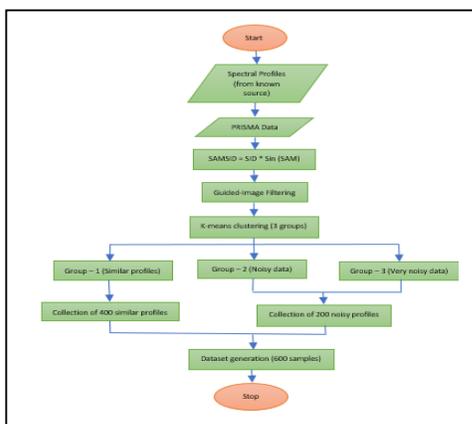The flowchart shown in Fig.1. explains the procedure followed for dataset preparation.

Fig. 1. Procedure followed to generate dataset

The dataset was created for binary classification, so this dataset had around 600 samples in which 60% were pure samples whereas remaining were noisy samples

3.6. Specification of the SVM learning algorithm

Dataset was further divided into training, testing and validation sets with 70%, 15% and 15% respectively. The learning algorithms details are: Radial Basis Function as kernel, sigma of range [1e-5 1e+5] and Box constraint of range [1e- 1e+5] were considered to find out the optimal value by Bayesian optimization. K-fold cross validation was performed with value of k as 10 to fit model with less error. For every class, kfloss of less than 5% was obtained.

3.7. Iso-bioclimatic map based classification

Bioclimatology deals with the interrelation between living organisms and climate studies, in particularly, plants and plant communities, considering the main climate variables that are relevant for species distribution. Usually for hydrology and ecology modelling, spatial information about main climatic variables that are relevant for species distribution will be used. In ecological studies, climate is considered as main factor that influences biological systems, affecting both the spatial distribution of living organisms and ecosystem function and processes, for example, wildfire regime or biomass production. Because of that, in hierarchical land classification

and mapping, bioclimate classification is considered a mandatory environmental layer that have be considered in analysis. In our case, iso-bioclimatic map obtained from ARPAS (Agenzia Regionale per la Protezione dell'Ambiente della Sardegna) was used to classify the mixed pixels left after the classification using SVM. That map comprises of 43 iso-bioclimatic zones with the combination of two macrobioclimates (Mediterranean pluviseasonal oceanic and Temperate oceanic), one variant of Temperate (Submediterranean), four classes of continentality (from weak semihyperoceanic to weak subcontinental), eight thermotypic horizons (from lower thermomediterranean to upper supratemperate) and seven ombrothermic horizons (from lower dry to lower hyperhumid) were identified and mapped [13]. Bacchetta et al.[14] has conducted a research activity on forest vegetation and serial-vegetation of Sardinia which was helpful to assign the suitable iso-bioclimate for each vegetation available in Sardinia. Different types of vegetation in accordance with iso-bioclimatic conditions were tabulated which is not shown here.

3.8. Generation of Fuel Map

Classified pixels were associated to the main vegetation fuel types of EFFIS/FUELMAP project [5][4]. Fuel weightage was assigned to the fuel types depending upon its proneness to fire in the range of 1 to 10. In this work, following the same principle, conversion from classified map to fuel map was done.

4. RESULTS AND DISCUSSION

As per the explained methodology, the PRISMA hyperspectral image was classified as shown in Fig. 2. Classified map was validated using different sources available from national or regional portals and Google Earth. The validation details are showed an overall accuracy of 86% and kappa coefficient of 0.79. Fig. 2. shows the image classified in to 14 different classes and labelled as 1. water, 2. bare soil, 3. buildings, 4. grasslands, 5. broad-leaved forest, 6. coniferous forest, 7. mixed forest, 8. small woody features, 9. sclerophyllous vegetation, 10. Olive groves, 11. transitional woodland shrubs, 12. vineyards, 13. fruit and berry plantations and 14. Cropland. Classes 15, 16 and 17 are remained as mixed pixels that are classified in to evergreen shrubs (*Juniperetum turbinatae*), eurasian shrubs (*Quercetum calliprini*) and tall trees with under bushes (O*leetum sylvestris*) using iso-bioclimatic map as explained in section 3.7.
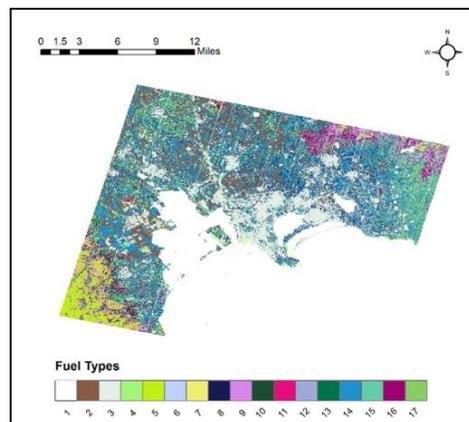

Fig. 2. Classified Map for southern part of Sardinia

5419

Performance metrics for each class was also identified to check the accuracy of the machine learning model. Performance metrics involves precision, recall and F1 score for analyzing. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it is better to look at both Precision and Recall.

TP, FP, FN stands for True Positive, False Positive and False Negative which could be obtained from the confusion matrix which is not shown. Table 1 shows the precision, recall and F1 score for each class. Almost every class has obtained value above 0.80 except vineyards for which suitable reference was not available to validate so it has obtained lesser but will carry out visual validation as soon as possible.

TABLE I.   PERFORMANCE METRICS

S.No	Precision	Recall	F1 Score
1	1	0.97	0.98
2	0.82	0.93	0.86
3	0.92	0.80	0.85
4	0.84	0.86	0.85
5	1	0.87	0.93
6	0.92	0.83	0.87
7	0.76	0.93	0.83
8	0.93	0.90	0.91
9	0.65	0.93	0.77
10	1	0.64	0.78
11	0.93	0.93	0.93
12	1	0.60	0.75
13	1	0.93	0.96
14	0.77	0.93	0.84
15	0.77	0.90	0.83
16	0.93	0.93	0.93
17	0.81	0.83	0.81

Fig. 3. shows the fuel map generated as per the methodology explained in the section 3.8. The fuel map has values ranging from 1 to 10 representing the fuel proneness to fire.
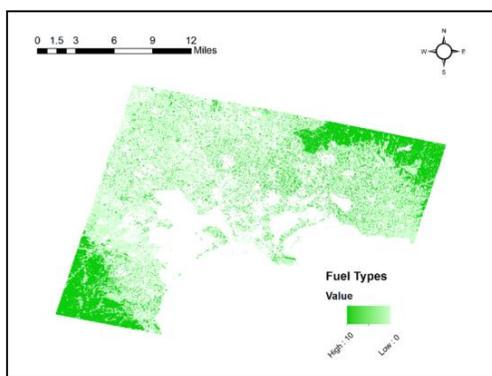

Fig. 3. Fuel Map for southern part of Sardinia

5. CONCLUSIONS

In this work, hyperspectral imagery from PRISMA, a satellite of Italian Space Agency is been used first time for the vegetation classification purpose. All the necessary fuel types in the southern part of Sardinian Island were detected. There is no dataset or literature available, so a new methodology is been proposed to generate samples for the fuel types. Support Vector Machine classifier was implemented to identity the fuel types with an overall accuracy of 86% and kappa coefficient of 0.79.

Mixed pixels were classified by associating vegetation types with iso-bioclimatic map and then the fuel map was generated by assigning the weights depending upon the fuel proneness to fire.

6. REFERENCES

[1] M. Niroumand-jadidi, F. Bovolo, and L. Bruzzone, "Water Quality Retrieval from PRISMA Hyperspectral Images : First Experience in a Turbid Lake and Comparison with Sentinel-2," 2020.

[2] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, no. xxxx, 2020.

[3] R. Archibald and G. Fann, "Feature selection and classification of hyperspectral images with support vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 674–677, 2007.

[4] *Advance EFFIS Report on Forest Fires in Europe , Middle East and North Africa 2018*. 2019.

[5] P. Toukiloglou, G. Eftychidis, I. Gitas, and M. Tompoulidou, "ArcFuel methodology for mapping forest fuels in Europe," *First Int. Conf. Remote Sens. Geoinf. Environ.*, vol. 8795, no. August, p. 87951J, 2013.

[6] Y. Du, C. Chang, and B. County, "New Hyperspectral Discrimination Measure for Spectral Characterization New hyperspectral discrimination measure for spectral characterization," no. August, 2004.

[7] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, 2013.

[8] T. Hien, T. Nguyen, D. Tai, D. Songsak, and S. Van Nam, "A method for k-means-like clustering of categorical data," *J. Ambient Intell. Humaniz. Comput.*, no. Berkhin 2002, 2019.

[9] P. J. Jones *et al.*, "FilterK: A new outlier detection method for k-means clustering of physical activity," *J. Biomed. Inform.*, vol. 104, no. October 2019, p. 103397, 2020.

[10] L. H. Thai, T. S. Hai, and N. T. Thuy, "Image Classification using Support Vector Machine and Artificial Neural Network," *Int. J. Inf. Technol. Comput. Sci.*, vol. 4, no. 5, pp. 32–38, 2012.

[11] M. Dalponte, H. O. Ørka, T. Gobakken, D. Gianelle, and E. Næsset, "Tree Species Classification in Boreal Forests With Hyperspectral Data," vol. 51, no. 5, pp. 2632–2645, 2013.

[12] M. Pinheiro, M. Zortea, D. Capella, Y. Edemir, C. Roberto, and D. S. Filho, "Remote Sensing of Environment Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data," *Remote Sens. Environ.*, vol. 179, pp. 66–78, 2016.

[13] S. Canu *et al.*, "Bioclimate map of Sardinia (Italy)," vol. 5647, 2015.

[14] G. Bacchetta, S. Bagella, E. Biondi, E. Farris, and D. Patterns, "Forest vegetation and serial vegetation of Sardinia (with map at the scale," no. September 2015, 2009.