Identification of pegmatite bodies, at a province scale, using machine learning algorithms: preliminary results

Teodoro, Ana Claudia, Santos, Douglas, Cardoso-Fernandes, Joana, Lima, Alexandre, Brönner, Marco

Ana Claudia M. Teodoro, Douglas Santos, Joana Cardoso-Fernandes, Alexandre Lima, Marco Brönner, "Identification of pegmatite bodies, at a province scale, using machine learning algorithms: preliminary results," Proc. SPIE 11863, Earth Resources and Environmental Remote Sensing/GIS Applications XII, 1186308 (12 September 2021); doi: 10.1117/12.2599600

SPIE.

Event: SPIE Remote Sensing, 2021, Online Only

Identification of pegmatite bodies, at a province scale, using machine learning algorithms: preliminary results

Teodoro, A.C.*[1,2], Santos, D.[1,2], J. Cardoso-Fernandes[1,2], Lima, A.[1,2], Brönner, M.[3]

[1]Institute of Earth Sciences, FCUP pole, Porto, Portugal; [2]Department of Geosciences, Environment and Spatial Planning Faculty of Sciences of the University of Porto, Portugal; [3]Geological Survey of Norway (NGU), Norway; *amteodor@fc.up.pt

ABSTRACT

Several raw materials for "green" energy production, such as high purity quartz, lithium, rare earth elements, beryllium, tantalum, and caesium, can be sourced from a rock type known as pegmatite. The GREENPEG project (https://www.greenpeg.eu/), started in May 2020, is developing and testing new and advanced exploration technologies and algorithms to be integrated and upscaled into flexible, ready-to-use economically efficient and sustainable methods for finding buried pegmatites and their "green" technology raw materials. One of the tasks of this project aims to apply different image processing techniques to different satellite images (Landsat, ASTER, and Sentinel-2) in order to automatically identify pegmatite bodies. In this work, we will present the preliminary results, regarding the application of machine learning algorithms (ML), more specifically, random forests (RF) and support vector machines (SVM) to one of the study areas of the project in Tysfjord, northern Norway, to identify pegmatite bodies. To be able to determine the classes that would make up the study area, geological data of the region, such as lithological maps, aeromagnetic data, and high-resolution aerial photographs, were used to define the four classes (1. pegmatites, 2. water bodies, 3. vegetation, 4. granite). All training locations were randomly selected, with 25% of the samples split into testing, and the remaining 75% split for training. The SVM algorithm presented more promising results in relation to overfitting and final image classification than RF. Testing the algorithms with several variables of parameters was able to make the process more efficient.

Keywords: Sentinel-2, pegmatites, random forests, support vector machines

1. INTRODUCTION

Due to the targets set by the European Climate Foundation for decarbonization by 2050, increasing demand for several raw materials (SRM) related to the production of "green" energy is expected. These minerals can be obtained from pegmatites - a rock enriched with rare elements found in low abundances such as Y, Th, U, Li and other metals like rare earth elements (REE)[1,2,3,4]. In order to meet this demand and thus ensure the future of Europe's technological development in "green" energy, the Greenpeg project aims to develop new and advanced cost-effective and sustainable technologies to find buried pegmatites and their "green"-tech raw materials. As test sites for the development of such a toolbox, three known pegmatite deposits with different types of pegmatites and in different settings were selected: Wolfsberg (Austria), Leinster (Ireland), and Tysfjord (Norway). With the significant improvement in coverage and accuracy of satellite data during the last decade, the use of remote sensing based on particularly adjusted satellite image processing became an important technology to locate and highlight possible areas of interest for prospecting. The methods applied comprise traditional image processing methods (e.g., RGB combinations, Band Ratios (BR) and Principal Component Analyses (PCA)), to support the main objective of this work, which is, to apply robust machine learning (ML) algorithms, such as Random Forest (RF) and Support Vector Machine (SVM) on images from the Sentinel-2 satellite, to automatically identify pegmatite bodies at a study area, with approximately 559 km², located in northern Norway. The images selected were those that presented less vegetation coverage, less cloud cover, and less

Earth Resources and Environmental Remote Sensing/GIS Applications XII,
edited by Karsten Schulz, Proc. of SPIE Vol. 11863, 1186308 · © 2021 SPIE
CCC code: 0277-786X/21/$21 · doi: 10.1117/12.2599600

snow, therefore, Summer images were chosen. The preliminary results of these two algorithms were compared and after analysis, new possible areas of interest for pegmatite prospecting were drawn. These preliminary results, besides being encouraging, are of great importance for the GREENPEG project, and can be used in technical reports for decision making about the location and prospection of pegmatites of economic interest.

2. STUDY AREA

Initially, the study area corresponded to the Tysfjord province with 10,913 km², but to improve the results and reduce the overfitting, the study area was reduced to the Tysfjord district, an area of approximately 559 km² located in northern Norway (Figure 1). The pegmatites of this area were classified such as NYF (Niobium, Yttrium and Fluorine) and the main minerals of these pegmatites are columbite-(Fe), allanite-(Ce), fergusonite-(Y), beryl, sulfides, and fluorite[5]. One of the most prominent pegmatites in the study area, the Håkonhals pegmatite (Figure 1) is about 25 m thick vertically and can reach 400 m in length and 200 m in width. All pegmatites from this region have sub-horizontal zones in rare lensoidal layers and minerals such as oxides, silicates, fluorite phosphates, and carbonates[1,6].

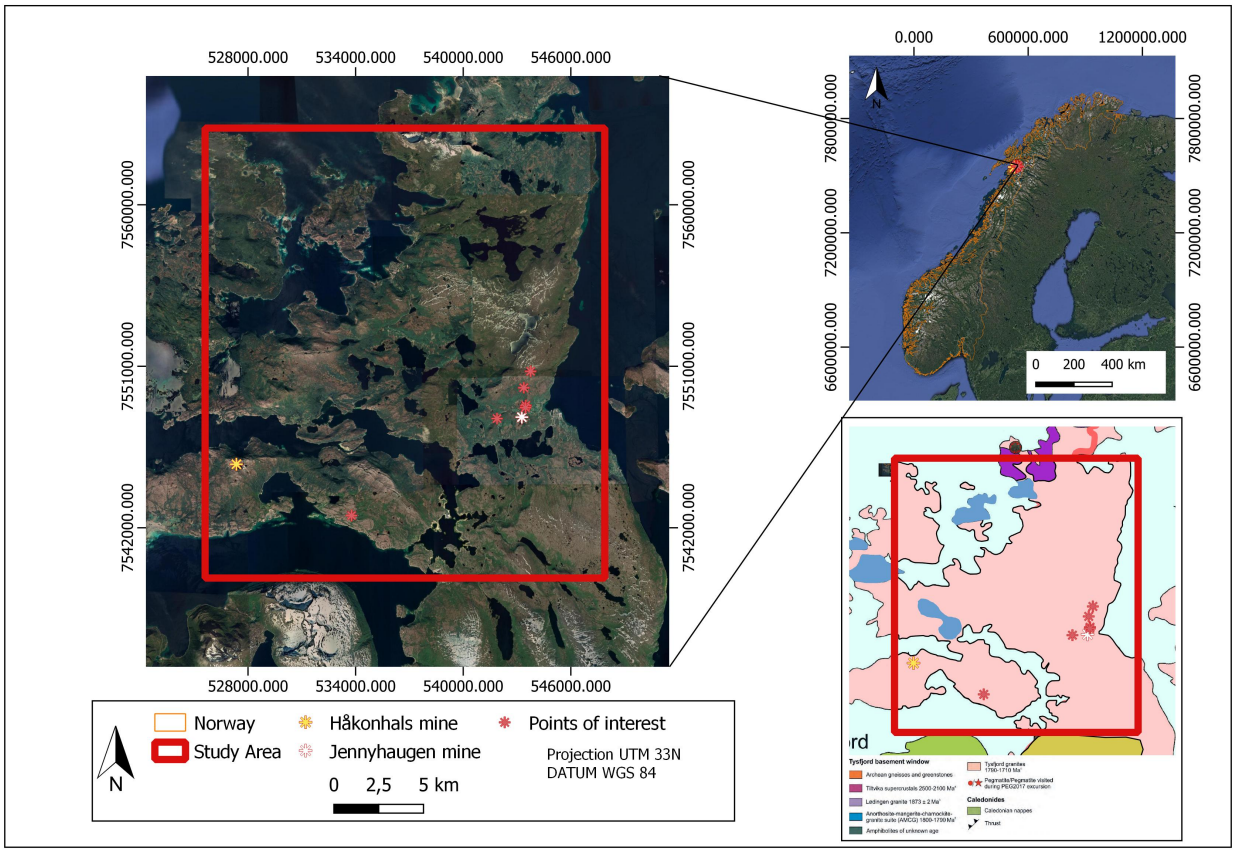

Figure 1. Study area location in Tysfjord district, Norway. The aerial photograph to the left shows the jagged landscape with high topography, intermitted by a large number of waterbodies from lakes and fjords. In the lower right corner, an adapted geological map5 of the district marks the most prominent known pegmatite deposits, where Håkonhals (6) and Jennyhaugen (5) mines are marked with yellow and white stars respectively.

3. METHODOLOGY

The ML algorithms were optimized and evaluated in Python programming language using an open-access library for ML called scikit-learn[7]. Due to its high-level interactive nature and the high availability of scientific libraries, the Python programming language is a powerful choice for algorithmic development and data analysis [7,8,9] . In this process, several models were created with different variables that correspond to intrinsically parameters linked to the Python code of the

algorithm, such as: (i) class imbalance, (ii) class balanced, (iii) data splitting considering the regions of interest (ROIs) as proposed by Cardoso-Fernandes et al[10] and; (iv) data splitting without considering the ROIs. The best of these models was then chosen.

3.1 Random Forest and Support Vector Machine Algorithms

Two algorithms were tested in this study: RF and SVM. The RF is a supervised classification algorithm that is robust for the classification of remote sensing data from multiple sources and much easier to apply than boosting-based methods and conventional bagging with a comparable tree-type classifier, capable of improving classification accuracy[11]. It is defined as a combination of independent tree predictors, where each tree depends on the values of a random vector with the same distribution and casts a unit vote for the most popular class in the input[12]. To ensure the randomness of the trees in the ensemble, a different subset of multispectral bands can be selected for use when splitting each node during the tree construction [13]. Taking this into account, two main hyperparameters need to be optimized: (i) the number of trees in the ensemble and; (ii) the number of features used in each node for splitting[14,15].

The SVM is also a ML algorithm proposed by Vapnik[16], and based on a statistical learning theory[17,15]. According to Cardoso-Fernandes et al[15], the SVM technique separates the dataset into a defined number of classes by trying to find an optimal hyperplane, which is a threshold that maximizes the distance of the margin between the class bounding or supporting hyperplanes. Noi and Kappas[14] concluded that the SVM classifier produces, on average, the largest overall accuracy and has less sensitivity to the size of the training samples. As pegmatites typically form small bodies, which consequently generate smaller training samples, SVMs are a powerful resource for classifying these areas.

3.2 Data Used and Image Acquisition

The GIS (Geographic Information System) data used in this study was, in part, provided by the GREENPEG project partners. These data were analyzed and the more appropriate data were selected and used. For example, among the data provided, the following were used in this study: regional geological data, such as geological maps[5], aeromagnetic data[18], and shp files for the study area, such as the limits of the areas of interest (province, district and prospecting scale), known pegmatite points and other shp file of interest to the study. In the shp file of pegmatites location, for example, only nine from thirty known pegmatite points were selected for this study. This happens due to many of the points were not picked up by the spatial resolution of the satellite data.

The Sentinel-2 Multispectral Instrument (MSI) has 13 spectral bands: 10 bands in the visible and near-infrared (VNIR) region and three on the shortwave infrared (SWIR)[19] region. The images downloaded have a cloud cover of less than 10%. The images were pre-processed using the Semi-Automatic Classification Plugin (SCP) plugin (version 7.0.15), available in the QGIS software (version 3.10.12). The atmospheric correction method used was the Dark Object Subtraction (DOS1)[20]. To select the best image for the study area, the Normalized Difference Vegetation Index (NDVI) was used in pre-processing step in order to analyze the vegetation and to select the image with less vegetation cover. Due to the snow cover in the study area, this factor was also put into consideration, and, to evaluate this, the Normalized Difference Snow Index (NDSI) was also computed and evaluated. Considering all these factors, the selected image for this preliminary study was from 28/09/2019 following the methodology developed by Cardoso-Fernandes et al. [21] and implemented by Santos et al. [22] in Brazil. The five bands from the Sentinel-2 image that are more suitable for this purpose were chosen (bands 2, 3, 4, 8, and 12).

3.3 Image Processing

After the selection and pre-processing of the satellite image, several traditional remote sensing methods were applied. These traditional methods were already applied in Portugal, Spain[21] and in southeastern Brazil[22] achieving encouraging results. The traditional methods employed were: (i) RGB combination: this method consists of combining different bands to produce different color combinations and to highlight geological units on a regional scale [21,22,23]; (ii) Band Ratios, which consists of a division of two bands to highlight specific spectral differences [21,22,24] and; (iii) PCA[25] (Principal Components Analysis), which is a multivariate statistical technique used to enhance and separate certain types of spectral signatures from the background. The best results were selected taking into consideration the effectiveness in highlighting areas of known pegmatites bodies and the less number of false positives. Only the best results (Table 1), were selected to be used in the next steps.

Table 1. The best traditional methods performed for Sentinel-2 images. In the left column is presented the method applied and in the right column, the Sentinel-2 bands that were used.

Method	Bands used
RGB combination	3-2-12
Band Ratios	3/8 4/12
PCA	3,8

3.4 Class Definition

After the selection of the best results considering the traditional image processing methods, the next step was to use these data together with geological maps[5], aeromagnetic data[18], high-resolution aerial photographs[26], and NDVI to understand and identify the elements present in the study area and, therefore, to identify the classes. Some authors[27, 28, 29] argue that each class should have at least between 30 and 60 samples. On another hand, Eastman[30], suggests that the number of pixels in each training set should not be less than ten times the number of bands. Respecting the suggested parameters described above, first, we defined six classes (Pegmatite bodies, Caledonian nappes, Granite, Ice/snow, Water, and Vegetation), but after decreasing the size of the study area to the Tysfjord district, the number of classes was reduced to four (Pegmatite bodies, Granite, Water, and Vegetation). Table 2 presents the relationship between the number of pixels and samples per class.

Table 2. Number of polygons (samples) and pixels for each study class.

Class	Polygon per class	Pixel per class
Pegmatites	30	79
Granite	30	356
Water	31	181
Vegetation	39	182

As can be observed in Table 2, the number of pixels for each class is more than 10 times the number of bands (five bands) and all classes have at least 30 sampling polygons.

3.5 Class Separability

In order to check if the classes are well separated and distinguished, and thus, preventing pixels from being misclassified, the class separability was analyzed using the PCI Geomatics software. If one or more classes present low separability, the class identification is redone. The spectral proximity between granite and pegmatites classes caused confusion in the classified image. Tables 3 and 4 show the class separability for the training samples.

Table 3. Class separability for the 1st sampling without parameters specified.

	Pegmatites	Granite	Water	Vegetation
Granite	1.048349			
Water	1.999746	1.999914		
Vegetation	1.822837	1.956871	1.967556	

Table 4. Class separability for the 2nd sampling without parameters specified.

	Pegmatites	Granite	Water	Vegetation
Granite	1.238486			

Water	1.999746	1.999952	
Vegetation	1.923277	1.992713	1.986382

This first training group (first and second sampling) was made without following the parameters specified in section 1.3. Tables 5 and 6 show the class separability after employing the parameters suggested in section 1.3, where it is possible to see that the less separability corresponds to pegmatites and granite classes.

Table 5. Class separability for 3rd sampling following the suggested parameters.

	Pegmatites	Granite	Water	Vegetation
Granite	1.681762			
Water	2.000000	2.000000		
Vegetation	1.994374	2.000000	2.000000	

Table 6. Class separability for 5th sampling following the suggested parameters.

	Pegmatites	Granite	Water	Vegetation
Granite	1.681132			
Water	2.000000	2.000000		
Vegetation	2.000000	2.000000	2.000000	

As it is possible to check in Tables 5 and 6, it is clear that the separability result is better after following the suggested parameters. The smaller class separability that involved our target sample (pegmatites) was between pegmatites and granite (1.681132) for both samplings. Although this result does not present a very good value by itself, this is a good improvement when compared to sampling without following the suggested parameters (Table 3 and 4). All the other classes presented maximum separability.

3.6 Data split into training and tests subsets

The dataset was split, considering 25% of the polygons for testing, and the remaining 75% for training. The data split was made according to their ROIs as proposed by Cardoso-Fernandes et al.[10], to ensure the maximum independence between the training and test subsets.

3.7 Model creation

The first step in the model creation was to define which parameters to optimize and the corresponding parameter range (or variation) to be used in the automatic search (grid-search). For RF, a preliminary stage was used in which only the number of trees, defined by the function 'n_estimators', was tested. In this study, the number of trees tested ranged from 10 to 500, with a constant increase after n_estimators = 50. For SVMs, the initial parameters start at 0.001 and are multiplied by 10 until they reach 100 (logarithmic scale). The defined parameter range was tested in a grid-search with cross-validation (stratified 5-fold) to choose the optimal parameters to be used in the Tysfjord demonstration site. For Sentinel-2 data with the RF algorithm, the optimal number of trees was 50[10].

3.8 Model evaluation

At this step, the best model returned by the grid search (in the model creation step) was evaluated using the test subset. After evaluating the model performance on the test set and achieving acceptable results, it was possible to proceed to the classification process. Different evaluation metrics can be employed, including the overall accuracy (OA) and/or kappa statistics. Moreover, a confusion matrix can be obtained considering the test subset, as well as a classification report containing the precision, recall, and f1-score for each class[10].

4. PRELIMINARY RESULTS AND DISCUSSION

As we can see in Table 7, both algorithms have very high values for the training scores, test scores, and OA. However, these high values indicate overfitting of the models. Comparatively, we can see that SVM performs better in relation to overfitting, despite having a similar value for accuracy (0.963 for SVM and 0.961 for RF). The SVM presents more acceptable values for test and train scores when compared to RF.

Table 7. Classification results summary for both algorithms (RF and SVM).

	SVM Results	RF Results
Mean train score	0.963	0.997
Mean test score	0.957	0.960
Overall accuracy	0.963	0.961
Kappa hat	0.989	0.910

As for classification, both algorithms had similar results and were able to highlight four of the seven known pegmatite areas in the Tysfjord district. As we can see in more detail in Figure 2, comparing the two algorithms, we can say that the RF algorithm classifies a greater amount of pixels as granite while the SVM algorithm highlighted more pixels as water bodies and vegetation.

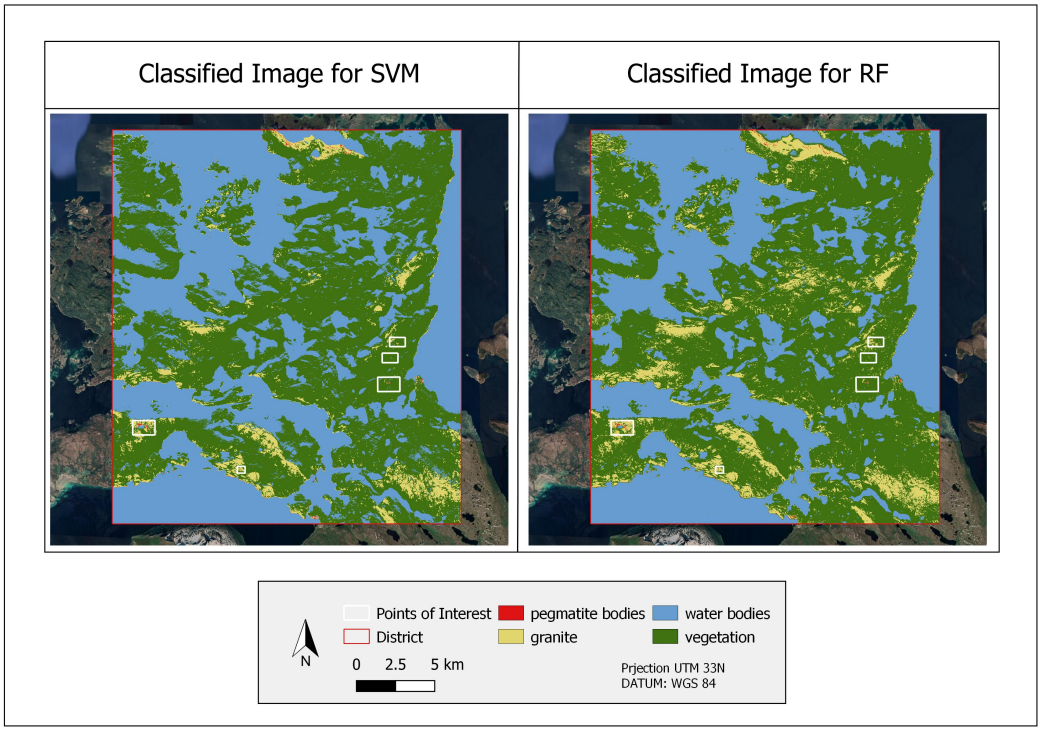

Figure 2. Comparison of results for the SVM (left) and RF (right) algorithm classifications.

When comparing the two algorithms, it is noticeable that RF performed better for the granite outcrop classification, identifying and highlighting the pixels of this class more accurately. On the other hand, RF proved to be less sensitive to spectral differences between the classes, assuming and misclassifying pixels as exemplified in Figure 3, where it classified vegetation where it should be water. The SVM algorithm was more sensitive to these spectral differences, resulting in a more assertive classification and a better separation between the classes, as can be seen in Figure 4. However, it did not work well for the classification of the granite outcrops, highlighting most of the exposed granite areas as vegetation.

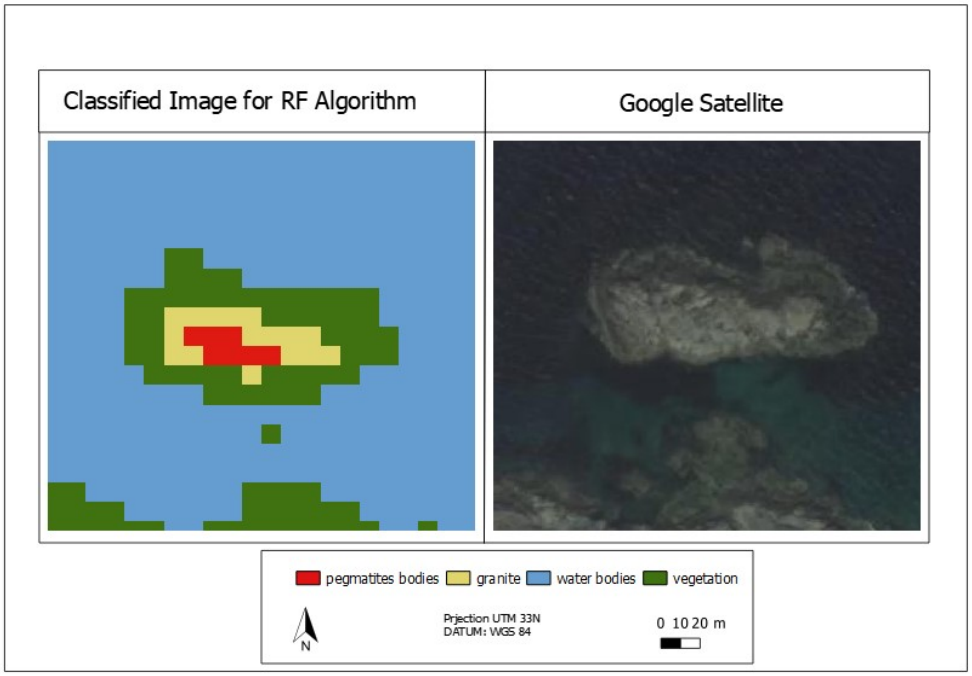

Figure 3. Comparison between RF classification and Google satellite image where RF assumes as vegetation pixels that correspond to water.

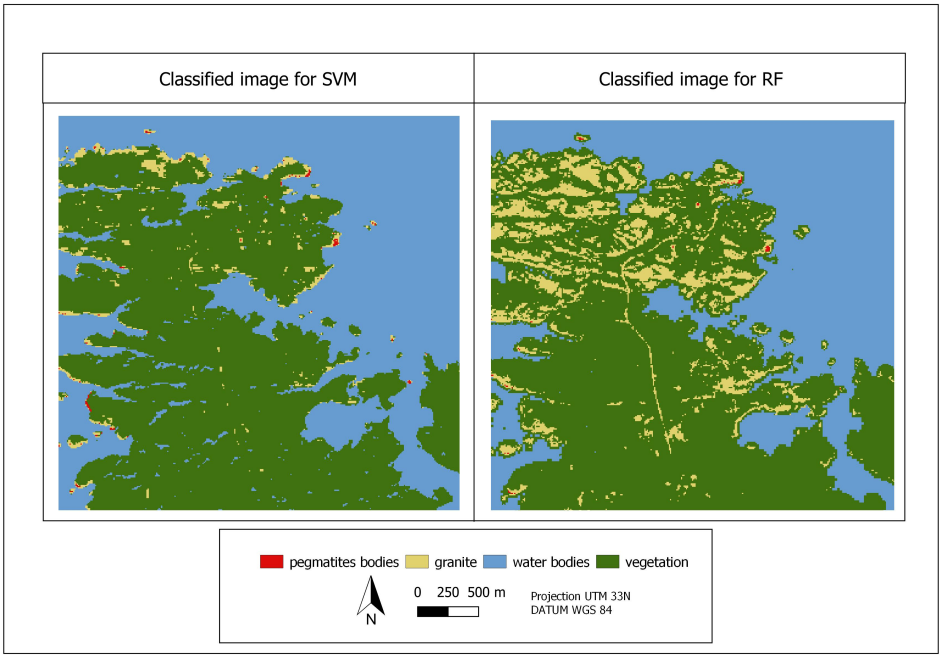

Figure 4. Exemplification of the difference in granite classification between the two algorithms.

Both methods obtained a similar result regarding the classification of our study target, highlighting, besides the already known pegmatite areas, new possible areas of interest for prospecting. Figure 5, compares the classification of the two algorithms for the areas of already known pegmatites.

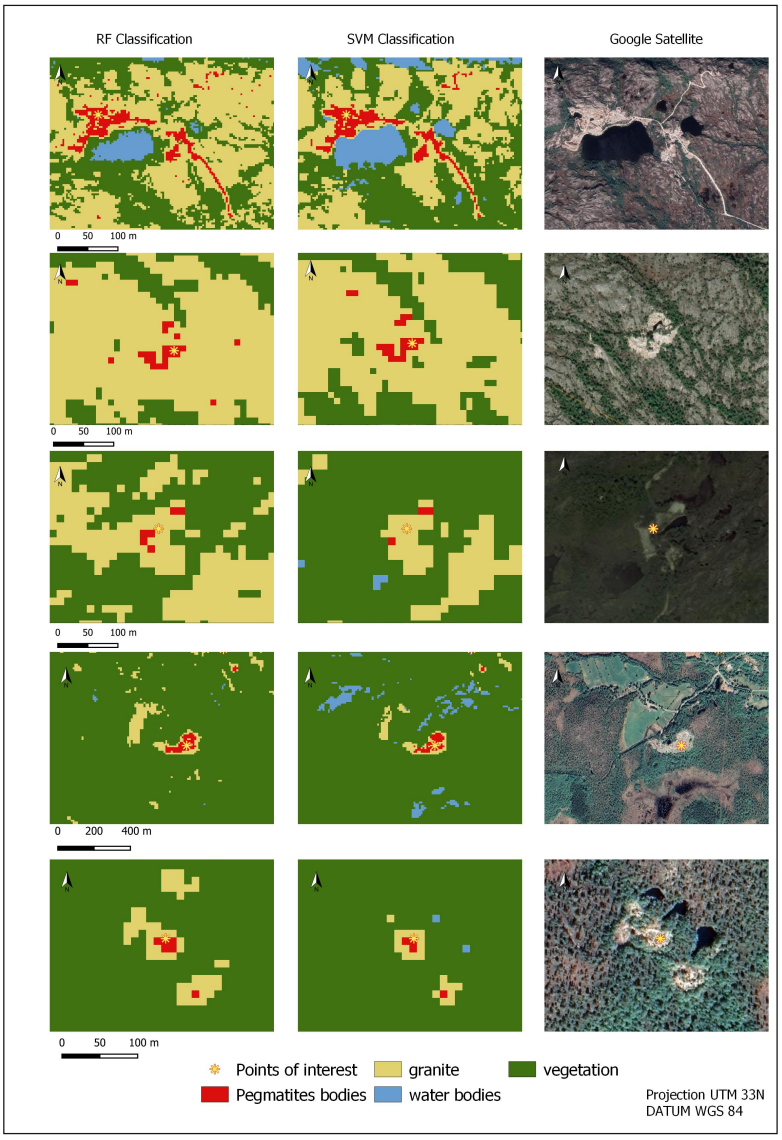

Figure 5. Comparison of the classification of the two algorithms for the known pegmatite areas.

Based on the patterns in which the algorithms highlight areas with already known pegmatites, new possible areas of interest were selected (Figure 6), and compared with the high-resolution and Google Earth images to identify false positives. In total four areas were selected: point a) located at coordinates 536401.86 m E/ 7556336.90 m N, having RF highlighted a larger area as pegmatites (red color) than SVM; point b) located at coordinates 528301.13 m E/ 7559019. 79 m N, corresponding to an area on the edge of a dirt road, that unlike the previous one, presented a higher number of highlighted pixels in the SVM classification; point c) located at coordinates 543539.60 m E / 7548600.89 m N, close to three points of pegmatites already known, which may indicate a possible deposit of tailings of these mines, with a higher number of pixels classified as pegmatite with the RF algorithm; and point d) at coordinates 527454.93 m E/ 7556955.90 m N, having obtained more prominence (larger number of pixels) with the SVM algorithm.

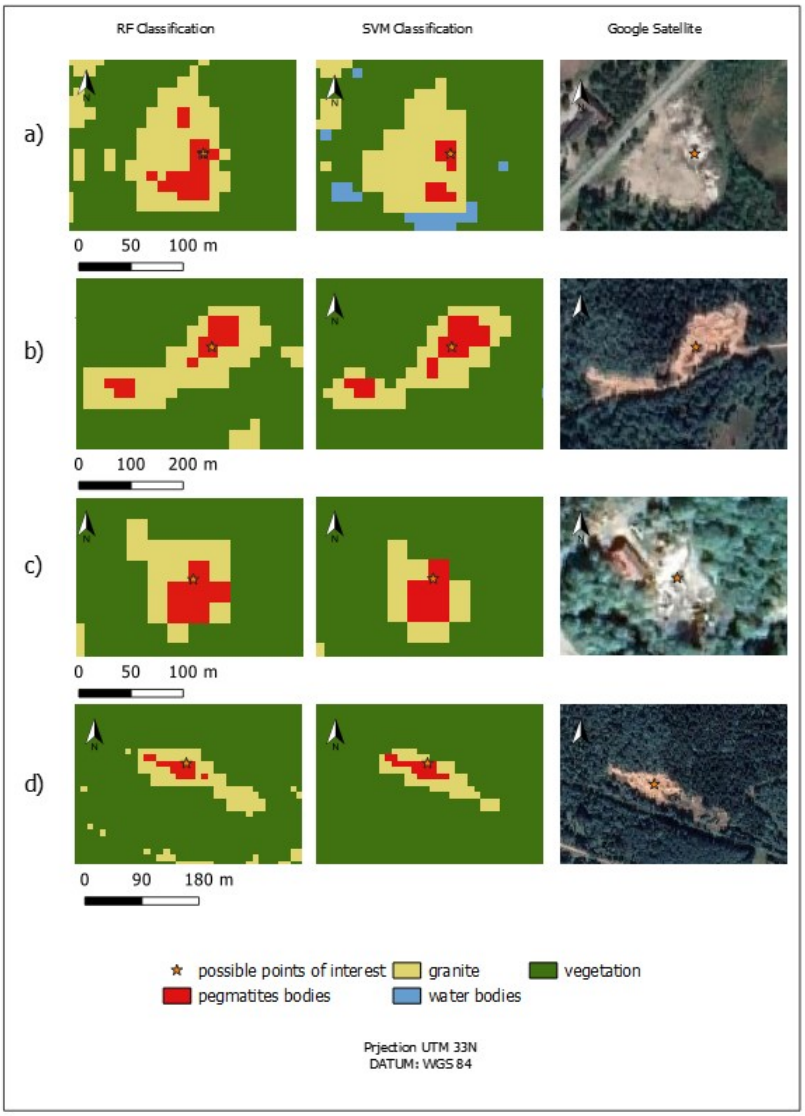

Figure 6. Comparison of the prominence of possible areas with the presence of pegmatites.

Point a) is close to the city of Ulvåg, built on a large river bank. The material of this riverbank is extensively quarried just NE of this site. The material of the quarry is described by NGU (Geological Survey of Norway)[31] as mostly short-transported weathered material from local bedrock and therefore likely to be rich in feldspar and quartz from the surrounding granite. For Northern Norway, the grussy type of weathered bedrock is most characteristic with a low degree of chemical weathering and preserving e.g. feldspar minerals. The disintegration of former pegmatites due to weathering processes explains why the algorithm can give a positive answer.

For site b), although not confirmed yet, a similar setting seems likely. Pictures from Google street view confirm a quarry for grussy to sandy material, popular for locals to built and maintain local gravel roads. Core stones (Figure 7) indicate also here bedrock weathering.

Point c) is in the vicinity of Jennyhaugen, a former pit for quartz pegmatites and right next to the quartz pegmatite mine Nedre Øyvollen, which was in production by Quartz Corp until 2018. Some of the produced quartz was deposited on this site for unknown reasons and even though this is not an *in situ* deposit, the algorithm recognized the pegmatite.

Site d) is on Tranøy again, at the northern shore of the Brennvik lake and east of Vassbotn. The site is not described in NGU's archives or registered as an active quarry. However, from aerial pictures it shows a similar characteristic as site b). Furthermore, historical mines or quarries for feldspar pegmatites are nearby at Tommeråsen at the southern shore of Brennvik lake and underline a similar setting with potential weathered pegmatitic bedrock.

Figure 7: Picture of the quarry at Tranøy shows sandy to grussy type material. Core stones indicate potential bedrock weathering, which is common in this area and possibly even *in-situ*.

In this analysis, it was possible to verify that there was signal confusion between pegmatites with buildings and coastal areas such as intertidal/supratidal areas and, to a lesser degree, confusion with steep slopes thus generating false positives. Figure 8 gives examples of these false positives.

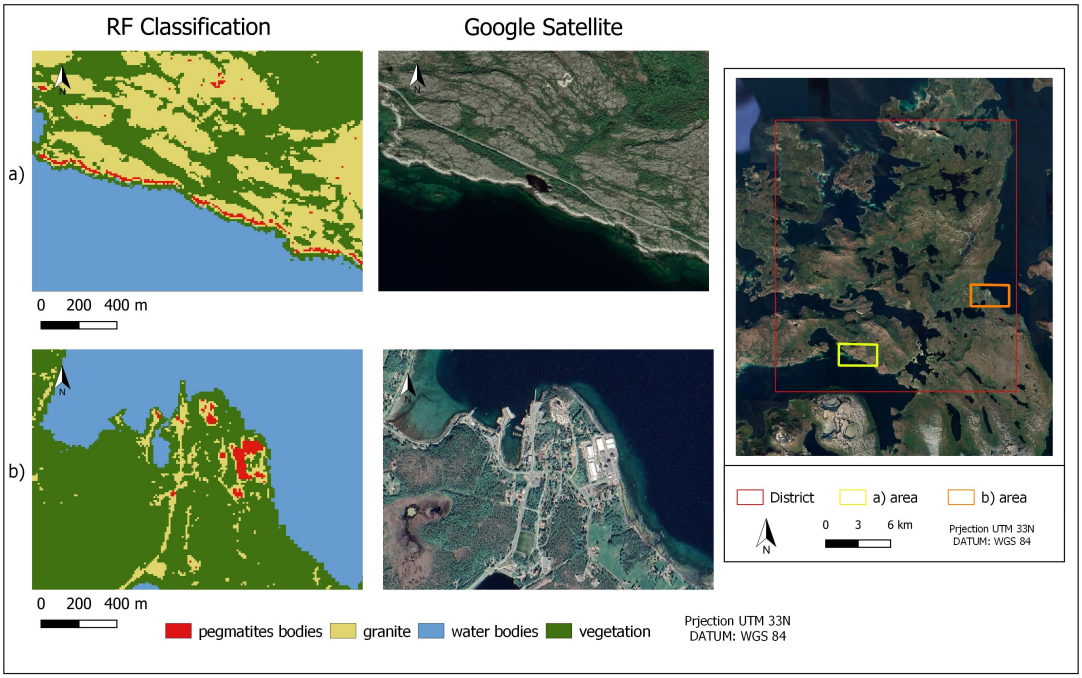

Figure 8. Examples of false positives obtained with the RF algorithm: a) an example of the signal confusion in a coastal area; b) an example of the signal confusion between buildings and pegmatites.

Although RF highlighted more pixels in areas of known pegmatites, the difference is so small that we can basically consider that the two algorithms achieved an equivalent pegmatite classification. For the other classes, RF classified the granite class better, while SVM was more sensitive to water and vegetation classes, classifying these two classes more accurately.

Overfitting and false positives proved to be the major obstacles in this study. The near-perfect score results in the algorithm training, indicate overfitting of the model and, consequently, misclassification of the image is expected. According to Cardoso-Fernandes et al[10,32], this happens because when the model is applied to the whole image, the performance score drops dramatically due to the inability of the algorithms to generalize to the new data introduced, thus resulting in misclassification. Of the several attempts to decrease the overfitting of the model, the one that worked best was to decrease the size of study area to be classified. With a smaller area, the results of the SVM algorithm improved, but overfitting was still noticed. On the other hand, even with a small improvement, the RF algorithm still has very high score values, and this is directly reflected in the image classification, which ended up misclassifying vegetation and water areas. The SVM algorithm, obtained a poor classification of granite class, which may also be a consequence of model overfitting.

5. CONCLUSIONS

This study proves that ML algorithms can be a very good technique for pegmatite localization because besides being able to identify very well-known pegmatite areas, are also able to highlight possible areas with the presence of pegmatites, which makes it possible to analyze new possible areas for prospecting. In general, the SVM algorithm performed better both in terms of overfitting and image classification. The RF algorithm was clearly affected by the high overfitting, resulting in a misclassification between the vegetation and water classes. It is noteworthy, that even with the overfitting present in both algorithms, the number of false positives was decreasing as the algorithm was improved. It is also important to mention that the classification generated by the algorithms allows a clean analysis of the results without the need, in this case, of adding filters, masks, or image post-processing, as it happens with some traditional methods (RGB combination, BR and PCA). Concluding, the SVM algorithm on the variable "balanced" with the "region of interest (ROI)", obtained better performance, in general, in image classification. As for the classification of pegmatites, both algorithms had equivalent results. The results of this study are promising since the algorithms were able to identify areas with pegmatite bodies. However, especially for site b) and d) more investigation and sampling of the sites need to be done to confirm our hypothesis and to finetune the algorithm to potentially differentiate and identify also weathered pegmatites. Continuing to work on new methods to avoid overfitting and improve image classification is a challenge for the near future, as well as obtaining samples from areas with the possible presence of pegmatites, in order to validate the results obtained.

ACKNOWLEDGMENTS

This study is funded by European Union's Horizon 2020 innovation programme under grant agreement No 869274, project GREENPEG: New Exploration Tools for European Pegmatite Green-Tech Resources. The work was also supported by Portuguese National Funds through the FCT project UIDB/04683/2020 - ICT (Institute of Earth Sciences). Joana Cardoso-Fernandes is financially supported within the compass of a Ph.D. Thesis, ref. SFRH/BD/136108/2018, by national funds from MCTES through FCT, and co-financed by the European Social Fund (ESF) through POCH – Programa Operacional Capital Humano – and NORTE 2020 regional program.

REFERENCES

[1] Hetherington, C. J., Mailloux, G. A. and Miller, B. V., "A multi-mineral U-(Th)-Pb dating study of the Stetind pegmatite of the Tysfjord region, Norway, and implications for production of NYF-rare element pegmatites during orogenic collapse," Lithos **398**–**399**(January), 106257 (2021).

[2] D London., [Pegmatites], Mineral. Assoc. Canada, Québec, 347 p. (2008).

[3] Budzianowski, W. M. and Gomes, J. F. P., "Perspectives for low-carbon electricity production until 2030: Lessons learned from the comparison of local contexts in Poland and Portugal," Energy Sources, Part B Econ. Plan. Policy **11**(6), 534–541 (2016).
[4] European Climate Foundation, [Roadmap 2050: a practical guide to a prosperous, low-carbon Europe: technical and economic assesment], Volume I, 387 (2010).
[5] Hamarøy, T., Iveland, E., Müller, B. A., Husdal, T., Sunde, Ø., Friis, H., Andersen, T., Johansen, T. S., Werner, R., Thoresen, Ø. and Olerud, S., [Norwegian Pegmatites I : Tysfjord-Hamarøy, Evje-Iveland, Langesundsfjord], Geol. Soceity Norw., Trondheim, 122 p. (2017).
[6] Husdal, T., Müller, A., Olerud, S. and Thoresen, Ø., "1 . Pegmatites of the Tysfjord-Hammaroy area, northern Norway", [Norwegian Pegmatites I : Tysfjord-Hamarøy, Evje-Iveland, Langesundsfjord], Husås, A. M., Ed., Geol. Soceity Norw., Trondheim, 3-47 (2017).
[7] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., and Louppe, G. "Scikit-learn : Machine Learning in Python," J. Mach. Learn. Res., 12, 2825–2830 (2012).
[8] Oliphant, T. E., " Python for Scientific Computing " Computing in Science & Engineering, **9**(3), 10-21 (2007).
[9] Millman, K. J. and Aivazis, M., "Python for scientists and engineers," Comput. Sci. Eng. **13**(2), 9–12 (2011).
[10] Cardoso-Fernandes, J., Teodoro, A. C., Lima, A., and Roda-Robles, E. "Semi-Automatization of Support Vector Machines to Map Lithium (Li) Bearing Pegmatites", Remote Sens., **12**(14), 2319 (2020).
[11] Gislason, P. O., Benediktsson, J. A. and Sveinsson, J. R., "Random forests for land cover classification," Pattern Recognit. Lett. **27**(4), 294–300 (2006).
[12] Pavlov, Y. L., "Random forests," Mach. Learn., 1–122 (2019).
[13] Müller, A.C. and Guido, S., [Introduction to machine learning with Python: A guide for data Scientists], O' Reilly Media, Inc, Sebastopol, USA, 376 p. (2016).
[14] Noi, P. T., and Kappas, M., "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery", Sensors, **18**(1), 18 (2017). [15] Cardoso-Fernandes, J., Teodoro, A. C. M., Lima, A. and Roda-Robles, E., "Evaluating the performance of support vector machines (SVMs) and random forest (RF) in Li-pegmatite mapping: preliminary results," Proceedings SPIE Volume 11156, Earth Resources and Environmental Remote Sensing/GIS Applications X, 111560Q. DOI: 10.1117/12.2532577 (2019).
[16] Vapnik, V. N., [The Nature of Statistical Learning Theory], Springer-Verlag, New York (1995).
[17] Yu, L., Porwal, A., Holden, E. J. and Dentith, M. C., "Towards automatic lithological classification from remote sensing data using support vector machines," Comput. Geosci. **45**, 229–239 (2012).
[18] Olesen, O., Brö Nner, M., Ebbing, J., Gellein, J., Gernigon, L., Koziel, J., Lauritsen, T., Myklebust, R., Pascal, C., Sand, M., Solheim, D. and Usov, S., "New aeromagnetic and gravity compilations from Norway and adjacent areas: Methods and applications," Pet. Geol. Conf. Proc. **7**(0), 559–586 (2010).
[19] European Space Agency (ESA)., "User Guides: Sentinel-2 Spatial Resolution," <https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial> (9 November 2020).
[20] Chavez JR, P. S., "Image-Based Atmospheric Corrections - Revisited and Improved," Eng. Remote Sens. **62**, 1025–1036 (1996).
[21] Cardoso-Fernandes, J., Teodoro, A. C. and Lima, A., "Remote sensing data in lithium (Li) exploration: A new approach for the detection of Li-bearing pegmatites," Int. J. Appl. Earth Obs. Geoinf. **76**, 10–25 (2019).
[22] Santos, D., Teodoro, A. C. M., Lima, A. and Cardoso-Fernandes, J., "Remote sensing techniques to detect areas with potential for lithium exploration in Minas Gerais, Brazil," Proceedings SPIE Volume 11156, Earth Resources and Environmental Remote Sensing/GIS Applications X; 111561F. DOI: 10.1117/12.2532744 (2019).
[23] Pour, A. B. and Hashim, M., "Hydrothermal alteration mapping from Landsat-8 data, Sar Cheshmeh copper mining district, south-eastern Islamic Republic of Iran," J. Taibah Univ. Sci. **9**(2), 155–166 (2015).
[24] Simon, N.Email Author, Aziz Ali, C., Mohamed, K.R., Sharir, K., "Best Band Ratio Combinations for the Lithological Discrimination of the Dayang Bunting and Tuba Islands, Langkawi, Malaysia," Sains Malaysiana **45** (2016).
[25] Loughlint W.P., "Principal Component Analysis for Alteration Mapping," Photogramm. Eng. Remote Sens. **57**(9), 1663–1669 (1991).
[26] Norge i Bilder, NIBIO, S. kartverk., "Norge i bilder," <https://www.norgeibilder.no/> (2 March 2021).
[27] Van Genderen, J. L., Lock, B. F. and Vass, P. A., "Remote Sensing: Statistical Testing of Thematic Map Accuracy," Remote Sens. Environ. **7**(1), 3–14 (1978).

[28] Richards, J. A. and Jia, X., [Remote sensing digital image analysis: An introduction], 4th edition, Springer, Berlin, 454 p. (2006).
[29] Fitz, P. R., Vieira, J. C. and Soares, M. C., "Use Of Sampling Polygons In Supervisioned Classifications Of Satellite Images," Entre-Lugar **10**(19), 319–341 (2019).
[30] Eastman, J. R., [IDRISI for Windows: Introdução e exercícios tutoriais.], Centro de Recursos Idrisi Brasil – UFRGS (2018).
[31] Geological survey of Norway., "Database for industrial minerals," <https://geo.ngu.no/kart/mineralressurser_mobil/> (7 April 2021).
[32] Géron, A., [Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems], O'Reilly Media, Sebastopol, 568 p. (2017).