

MESTRADO EM SISTEMAS DE INFORMAÇÃO GEOGRÁFICA E ORDENAMENTO DE TERRITÓRIO

Remote sensing techniques for identification of NYF pegmatites in Tysfjord, Norway

Douglas Barbosa dos Santos

M

2021



Douglas Barbosa dos Santos

**Remote sensing techniques for the
identification of NYF pegmatites in Tysfjord,
Norway**

Dissertation carried out under the Master`s degree in Geographic Information Systems
and Spatial Planning, supervised by Professor Ana Cláudia Moreira Teodoro
and by Professor Alexandre Martins de Campos Lima

Faculdade de Letras da Universidade do Porto

2021

Douglas Barbosa dos Santos

Remote sensing techniques for the identification of NYF pegmatites in Tysfjord, Norway

Dissertação realizada no âmbito do Mestrado em Sistemas de Informações Geográficas
e Ordenamento do Território, orientada pela Professora Doutora Ana Cláudia Moreira
Teodoro.

e pelo Professor Doutor Alexandre Lima.

Membros do Júri

Professor Doutor (escreva o nome do/a Professor/a)

Faculdade (nome da faculdade) - Universidade (nome da universidade)

Professor Doutor (escreva o nome do/a Professor/a)

Faculdade (nome da faculdade) - Universidade (nome da universidade)

Professor Doutor (escreva o nome do/a Professor/a)

Faculdade (nome da faculdade) - Universidade (nome da universidade)

À minha família e amigos, que me incentivaram nos momentos difíceis e compreenderam a minha ausência enquanto eu me dedicava à realização deste trabalho.

Summary

Declaração de honra	5
Agradecimentos	6
Resumo.....	7
Abstract	8
List of Figures	9
1.Introduction	14
1.1. The Greenpeg Project.....	15
1.2. GIS (Geographic Information System) and Remote Sensing	16
1.2.1. The GIS	16
1.2.2. Remote sensing.....	18
1.3. Pegmatites	20
1.3.1. Homogeneous pegmatites.....	21
1.3.2. Zoned Pegmatites	21
1.3.3. Layered pegmatites	22
1.4. Study Area	22
1.4.1. The Jennyhaugen Pegmatite.....	23
1.4.2. The Nedre Øyvollen Pegmatite.....	24
The Håkonhals Pegmatite	24
1.4.3. 24	
1.5. Structure of the work	25
2.Metodology	27
2.1. Image Pre-processing	29
2.2. Image Selection and Indices	29
2.3. Spectral Library and Analysis.....	29
2.3.1. Spectral Library	29
2.3.2. Spectral Analyses	30
2.4. Image Processing.....	32
2.4.1. RGB combination	32
2.4.2. Band Ratios	32
2.4.3. Principal Components Analyses (PCA).....	33
2.5. Machine Learning (ML) Algorithms	34

2.5.1. The Random Forest Algorithm.....	35
2.5.2. The Support Vector Machine Algorithm.....	35
2.5.3. Reconnaissance of the area	36
2.5.4. Class Separability	37
2.5.5. Split the data into training and tests subsets	39
2.5.6. Model creation	39
2.5.7. Model evaluation.....	41
3.Results and Discussion	42
3.1. First set: Results for LCT traditional methods	42
3.1.1. RGB Combination.....	42
3.1.2. Band Ratios	43
3.1.3. Princpail Components Analyses (PCA).....	46
3.1. Second set: A new approach for traditional methods.	48
3.1.1. Principal Components Analyses.....	48
3.1.2. Band Ratios	51
3.2. Third set: Machine Learning algorithms.....	57
3.2.1. Results and comparasion.....	57
4.Conclusions	65
4.1. Future work	66
References.....	67
Attachment	78

Declaração de honra

Declaro que a presente dissertação é de minha autoria e não foi utilizado previamente noutro curso ou unidade curricular, desta ou de outra instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escrupulosamente as regras da atribuição, e encontram-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referência. Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

[Porto, Setembro de 2021]

[Douglas Barbosa dos Santos]

Agradecimentos

Reservo os primeiros agradecimentos dessa tese aos meus queridos orientadores, prof. Doutora Ana Cláudia M. Teodoro e Prof. Doutor Alexandre M. C. de Lima, que foram sempre repletos de sabedoria e paciência e por isso, os agradeço do fundo do meu coração não apenas pela orientação prestada, mas também pelas oportunidades que me foram confiadas, em especial a bolsa de investigação no projecto Greenpeg, sem a qual esse trabalho não teria sido desenvolvido.

Aos meus pais e irmã, por todo o suporte e apoio ao longo da minha caminhada. Sem vocês nada disso seria possível e espero um dia poder-lhes retribuir.

Aos familiares que sempre me desejaram bem, em especial a minha tia Marilucia Santos, por todas às vezes que me socorreu em momentos de dificuldade.

A minha melhor amiga, Bruna Bianca Brandalise Piva que, mesmo com a distância, nunca deixou de apoiar-me e ouvir em momentos de necessidade.

Aos professores, pelas correções e ensinamentos que me permitiram apresentar um melhor desempenho académico e profissional ao longo do curso.

Ao Prof. Doutor Alberto Gomes, por todas as dúvidas esclarecidas e orientações que guiaram o meu caminho e das quais eu nunca me esquecerei.

Ao Prof. Ricardo José Vieira Baptista, por todo apoio e ajuda recebida em momentos de necessidade.

Aos parceiros do projecto Greenpeg que, para além de ajudarem a validar resultados, também forneceram dados de extrema importância para a realização desse trabalho.

E por fim, aos meus colegas de curso por compartilharem comigo tantos momentos de descobertas e aprendizado e por todo o companheirismo ao longo deste percurso.

Resumo

Com o avanço tecnológico cada vez mais voltado as energias verdes, não é surpreendente que as matérias primas necessárias para desenvolver esse setor tecnológico sejam uma preocupação cada vez mais crescente entre os países que buscam poupar, ou aumentar as suas reservas de materiais considerados críticos para o seu desenvolvimento. Os fatores anteriormente citados, somados a um esforço da União Europeia em diminuir sua dependência em relação a matéria críticas e matérias de terras raras da China deram origem ao projeto Greenpeg, um projeto formado por mais 13 parceiros da União Europeia entre instituições académicas e do setor privado. Este projeto possui entre seus objetivos específicos, o desenvolvimento de novos métodos focados na identificação de pegmatitos em 3 áreas de estudo, sendo elas Tysfjord (Noruega) que corresponde à área que é objeto de estudo desta tese, Leinster (Irlanda do Sul) e Wolfsberg (Austria). É nesse âmbito que essa tese de mestrado se enquadra. Desenvolvida juntamente a este projeto, esta tese tem como principal objetivo a aplicação, desenvolvimento e avaliação de métodos de deteção remota em identificar pegmatitos do tipo NYF. Para isso, foram aplicados métodos tradicionais (composição RGB (Red, Green and Blue), Band Ratios (BR) e Análise de Componentes Principais(PCA)) já conhecidos na identificação de pegmatitos do tipo LCT (Lithium-Cesium and Tantalum), com o objetivo de atestar sua capacidade em identificar pegmatitos do NYF (niobium-yttrium-fluorine). Para além disso, também foi proposta uma nova abordagem onde os métodos tradicionais já conhecidos foram adaptados para as características dos pegmatitos de Tysfjord. Por último foram aplicados algoritmos machine learning (Random forest e Support Vector Machine) em dados do Sentinel-2, com o mesmo objetivo de identificar pegmatitos na área de estudo. Os resultados desta tese são promissores e fazem uma grande contribuição para a deteção remota aplicada para fins prospetivos.

Palavras-chave: Greenpeg, Deteção Remota, Materias primas criticas, elementos de terras raras, algoritmos de machine learning.

Abstract

With technological advances increasingly focused on green energy, it is no surprise that the raw materials needed to develop this technological sector are a growing concern among countries seeking to save or increase their reserves of materials considered critical to their development. These factors, combined with an effort by the European Union to reduce its dependence on critical materials and rare lands from China, gave rise to the Greenpeg project, a project formed by 13 partners from the European Union, including academic institutions and the private sector. This project has among its specific objectives, the development of new methods focused on the identification of pegmatites in 3 study areas, being Tysfjord (Norway) which corresponds to the area which is the object of study of this thesis, Leinster (South Ireland) and Wolfsberg (Austria). For this, traditional methods (RGB combination (Red, Green and Blue), Band Ratios (BR) and Principal Components Analysis (PCA)) already known in the identification of LCT (Lithium-Cesium and Tantalum) type pegmatites were applied, to attest their capacity in identifying NYF (niobium-yttrium-fluorine) pegmatites. Furthermore, a new approach was also proposed where the known traditional methods were adapted to the characteristics of the Tysfjord pegmatites. Finally, machine-learning algorithms (Random forest (RF) and Support Vector Machine(SVM)) were applied to Sentinel-2 satellite, with the same objective of identifying pegmatites in the study area. The results of this thesis are promising and make a great contribution to remote sensing applied for prospective purposes.

Keywords: Greenpeg, Remote sensing, Critical raw materials, Rare earth elements, Machine learning algorithms.

List of Figures

FIGURE 1. THE GIS OPERATIONS (SOURCE: MARTIN (1996)).	18
FIGURE 2. EXAMPLE OF THE FUNCTIONALITY OF PASSIVE AND ACTIVE SENSORS. (	19
FIGURE 3. EXEMPLE OF THE PROCESSES OF EMR INTERACTION WITH THE EARTH AND OBTAINING INFORMATION. (.....	19
FIGURE 4. REPRESENTATION OF THE ELECTROMAGNETIC SPECTRUM IN μM . (	20
FIGURE 5. INTERNAL STRUCTURE OF ZONED PEGMATITES IN HORIZONTAL SECTIONS (	22
FIGURE 6. LOCATION OF THE STUDY AREA, TYSFJORD, NORWAY. (SOURCE: MODIFIED FROM MÜLLER <i>ET AL.</i> (2021)).	23
FIGURE 7. HIGH RESOLUTION IMAGES OF THE HÅKONHALS AND JENNYHAUGEN MINES (	25
FIGURE 8. STEP BY STEP PROCESS FOR THE DEVELOPMENT OF A NEW APPROACH FOR IDENTIFICATION OF NYF PEGMATITES.....	28
FIGURE 9. SPECTRORADIOMETER USED TO COLLECT THE SPECTRAL DATA OF THE TYSFJORD SAMPLES.	30
FIGURE 10. SPECTRUM OF A SAMPLE OF AMAZONITE FROM TYSFJORD IN RELATION TO THE SENTINEL 2 SATELLITE BANDS.	31
FIGURE 11. WAVELENGTH FOR SENTINEL 2 (	32
FIGURE 12. WORKFLOW OF THE ML APPLICATION PROCEDURE.	34
FIGURE 13. BEST PARAMETERS ACHIEVED FOR THE RF ALGORITHM, USING SENTINEL-2.....	39
FIGURE 14. CROSS-VALIDATION SCORE AS A FUNCTION OF THE TESTED PARAMETER FOR SENTINEL-2.....	40
FIGURE 15. BEST PARAMETERS LOCATION IN A GRID SEARCH FOR THE SVM ALGORITHM.	40
FIGURE 16. CROSS-VALIDATION SCORE OBTAINED FOR EACH PARAMETER COMBINATION.	41
FIGURE 17. RGB COMPOSITION OF BANDS 3-2-12 FOR SENTINEL-2.	42
FIGURE 18. RGB COMBINATION 3-2-12 FOR SENTINEL-2. (.....	43
FIGURE 19. BAND RATIO 3/8 FOR SENTINEL-2.	44
FIGURE 20. BAND RATIOS 4/12 FOR SENTINEL-2.....	44
FIGURE 21. LCT PEGMATITE HIGHLIGHTED IN BRASIL. BR 3/8 FOR SENTINEL-2 BANDS. (	45
FIGURE 22. LCT PEGMATITE HIGHLIGHTED IN BRASIL. BR 4/12 FOR SENTINEL 2 BANDS. (	46
FIGURE 23. PC 2 BANDS OF THE BANDS 3,8 FOR SENTINEL-2.....	47
FIGURE 24. SELECTIVE PCA (2 PCs) FOR SENTINEL-2. (	48
FIGURE 25. PC 2 BANDS OF THE BANDS 2,8 FOR SENTINEL 2.....	49
FIGURE 26. COMPARISON OF THE RESULT BETWEEN PC 3,12 AND 2,12.....	50
FIGURE 27. PC 2 BANDS OF THE BANDS 8,12 FOR SENTINEL-2.....	51

FIGURE 28. BAND RATIOS 3/8 FOR SENTINEL-2. THE PIXELS OF THE MINES ARE HIGHLIGHTED IN LIGHT GREY..	52
FIGURE 29. BAND RATIOS 4/8 FOR SENTINEL-2. THE PIXELS OF THE MINES ARE HIGHLIGHTED IN LIGHT GREY..	53
FIGURE 30. BAND RATIOS 4/6 FOR SENTINEL-2. THE PIXELS OF THE MINES ARE HIGHLIGHTED IN LIGHT GREY..	54
FIGURE 31. COMBINATION OF RATIOS 4/8, 2/8 AND 8/12 IN RGB COMPOSITE	55
FIGURE 32. COMBINATION OF RATIOS 4/8, 3/8 AND 8/12 IN RGB COMPOSITE.	56
FIGURE 33. COMBINATION OF RATIOS 4/8, 8/4 AND 8/12 IN RGB COMPOSITE.	57
FIGURE 34. COMPARISON OF RESULTS FOR THE SVM (RIGHT) AND RF (LEFT) ALGORITHM CLASSIFICATIONS. ..	58
FIGURE 35. COMPARISON OF THE CLASSIFICATION OF THE TWO ALGORITHMS FOR THE KNOWN PEGMATITE AREAS.	59
FIGURE 36. EXEMPLIFICATION OF THE DIFFERENCE IN GRANITE CLASSIFICATION BETWEEN THE TWO ALGORITHMS.	60
FIGURE 37. COMPARISON BETWEEN RF CLASSIFICATION AND GOOGLE SATELLITE IMAGE WHERE RF ASSUMES AS VEGETATION PIXELS THAT CORRESPOND TO WATER.	61
FIGURE 38. COMPARISON OF THE PROMINENCE OF POSSIBLE AREAS WITH THE PRESENCE OF PEGMATITES	62
FIGURE 39. EXAMPLES OF FALSE POSITIVES OBTAINED WITH THE RF ALGORITHM. SOURCE: BY AUTHOR ADAPTED FROM USGS AND GOOGLE EARTH. A) AN EXAMPLE OF THE SIGNAL CONFUSION IN A COASTAL AREA; B) AN EXAMPLE OF THE SIGNAL CONFUSION BETWEEN BUILDINGS AND PEGMATITES.	63
FIGURE 40. COMPARISON OF THE BEST RESULTS OF EACH SET.....	64

LIST OF TABLES

TABLE 1. LIST OF PARTNERS IN THE GREENPEG PROJECT	16
TABLE 2. BAND RATIOS TESTED FOR THE SECOND SET, CONSIDERING SENTINEL-2 DATA.	33
TABLE 3. RELATIONSHIP BETWEEN THE NUMBER OF PIXELS AND SAMPLES PER CLASS	37
TABLE 4. CLASS SEPARABILITY FOR THE 1ST SAMPLING WITHOUT PARAMETERS SPECIFIED.....	38
TABLE 5. CLASS SEPARABILITY FOR THE 2ND SAMPLING WITHOUT PARAMETERS SPECIFIED.....	38
TABLE 6. CLASS SEPARABILITY FOR 3RD SAMPLING FOLLOWING THE SUGGESTED PARAMETERS.	38
TABLE 7. CLASS SEPARABILITY FOR 5TH SAMPLING FOLLOWING THE SUGGESTED PARAMETERS.	38
TABLE 8. CLASSIFICATION RESULTS SUMMARY FOR BOTH ALGORITHMS (RF AND SVM).	58

List of abbreviations and acronyms

CRM	CRITICAL RAW MATERIALS
REE	RARE EARTH ELEMENTS
LREE	LIGHT RARE EARTH ELEMENTS
HREE	HEAVY RARE EARTH ELEMENTS
WTO	WORLD TRADE ORGANIZATION
LCT	LITHIUM-CESIUM AND TANTALUM
NYF	NIOBIUM-YTTRIUM-FLUORINE
EM	ELECTROMAGNETIC
EMR	ELECTROMAGNETIC RADIATION
USA	UNITED STATES OF AMERICA
RGB	RED, GREEN AND BLUE
BR	BAND RATIOS
PCA	PRINCIPAL COMPONENTS ANALYSES
PC	PRINCIPAL COMPONENTS
ML	MACHINE LEARNING
SVM	SUPPORT VECTOR MACHINE
RF	RANDOM FOREST
GIS	GEOGRAPHIC INFORMATION SYSTEM
DN	DIGITAL NUMBERS
NDVI	NORMALIZED VEGETATION INDEX
NDSI	NORMALIZED SNOW INDEX
UTM	UNIVERSAL TRANSVERSE MERCATOR
CPU	CENTRAL PROCESSING UNIT
Nb	NIOBIUM
Y	YTTRIUM
F	FLUORINE
Be	BERYLLIUM

U GOLD
Li LITHIUM
Cs CAESIUM
TA TANTALUM

1. Introduction

In a world that is increasingly concerned about the possible depletion of its natural resources and that increasingly aims at the decarbonization of its sectors, the reduction of carbon emissions in the atmosphere and the reduction of its impact, i.e., that of human activity in nature, the growing need for new technologies that meet this "green" demand and that enable (Barroso, et al., 2010; Grübler and Nakićenović, 1996; Vacari *et al.*, 2018). With the development of these "green" technologies comes also the demand for elements that are raw materials necessary for their production and operation. Among these elements, we can find minerals that fall between the so-called CRM (Critical raw Materials) and REE (Rare Earth Elements), which increases the need to locate new deposits of these minerals, whose demand is growing dramatically (Gamaletsos et al., 2019; Haxel *et al.*, 2002).

Critical Raw Materials (CRM) are minerals considered crucial to the economic development of a country (Bobba *et al.*, 2020). The European Union (EU) considers CRMs as fundamental to its economic development, with great importance in industry, being part of supply chain stages and becoming essential to the European way of life (European Commission, 2014). Due to this importance, and with increasing pressure regarding the availability of these raw materials, the EU decided to conduct a study every three years to establish a list of raw materials that are critical for the block. By the year 2020, 30 CRMs were defined, including Beryllium, REEs, Niobium, and the most recent addition to the list, Lithium (Blengini, 2020).

Rare Earth Elements (REE) are defined as a family of 17 elements in the periodic table including, for example, the elements lanthanum, scandium and Yttrium (Massari and Ruberti, 2013). According to Massari and Ruberti (2013), except for promethium, all REEs can be found in nature and, contrary to what the name suggests, are not such rare elements. According to the same authors, this confusion happened because when REE were discovered they were thought to be present in small quantities in the Earth's crust. According to Haxel *et al.* (2002), the term "rare" for these elements turned out to be a historical error that persisted through time, because in fact the most abundant REE can be compared, in terms of concentration in the Earth's crust, to industrial metals such as copper and bronze. Depending on their location on the periodic table, REE can be separated into two types: (i) light REE (LREE), which have atomic number 57-62, and; (ii) heavy REE (HREE) which have an atomic number between 63-71 and are less common, in terms of extraction, when compared to LREE (Massari and Ruberti, 2013). In recent decades,

there has been a drastic increase in the demand for less abundant REE, this is mainly due to the technological applications of these elements that have come to multiply during this period.

When we analyze the possible lines of technological development of the future it is possible to predict an ever-increasing demand for CRMs and REEs, which makes it understandable that many countries have applied protectionist policies in order to protect their domestic industries. As a consequence of these policies there is a reduction in exports of these elements to the EU (Massari and Ruberti, 2013; Mateus, 2020). The impacts of shortages of elements considered essential are not new to Europe. The very term "criticality" regarding raw materials was first used during the First World War, when Germany secured control of the production of tungsten, a material needed to sustain the war effort, where it forced the other powers involved, to struggle to find alternative sources of this material (Chakhmouradian *et al.*, 2015; Rodríguez *et al.*, 2018).

In a more recent case China, which has large REE reserves, has made a tough cut to its REE export quotas, leading the USA (United States of America), the EU and Japan to bring the case to the World Trade Organization (WTO) (Rodríguez *et al.*, 2018). It is undeniable that China dominates REE and CRM imports (Blengini, 2020) and with technological development turning more and more to the use of green energy, due to EU's decarbonization goals by 2050, the concern of finding new deposits of critical materials is increasing. It is in this context that the Greenpeg project is inserted

1.1. The Greenpeg Project

The Greenpeg project, is an EU-funded project that aims to develop, test and apply new technologies and exploration algorithms that will be used in the identification of CRM-rich LCT and NYF pegmatites. The project has thirteen partners spread over eight countries, including eight private sector companies and five academic/research institutions (table 1), and has as main objectives:

I. Enhance European consulting expertise and the development of innovative, home-grown ("Made in Europe") exploration processes and technologies.

II. Improve the European Raw Material Information System for European pegmatite ore deposits and related commodities in Austria, Norway, Ireland, Finland, Portugal and Spain.

III. Reduce the need for and costs of invasive exploration and lessen the environmental impacts of mining.

IV. Make a substantial contribution to the application and implementation of European Policies and Regulations in the field of sustainable production of raw materials, meeting the ambitious 2030 energy and climate objectives, minimizing the environmental impacts and risks associated with CRM exploration and mining, and gaining the trust of EU citizens.

Table 1. List of partners in the Greenpeg project

Partner name	Short name	Country
Universitet i Oslo	UIO	Norway
University College Dublin	UCD	Ireland
Terratec Geophysical Services GMBH & Co KG	TERRA	Germany
Geokompetenzzentrum Freiberg e.V.	GKZ	Germany
Blackstairs Lithium Ltd	BLI	Ireland
Norges Geologiske Undersøkelse	NGU	Norway
IFU GmbH Privates Institut Für Umweltanalysen	IFU	Germany
Universidad Del Pais Vasco	UPV	Spain
The University of Exeter	UNEXE	United Kingdom
ECM Lithium AT GmbH	ECM	Austria
Universidade do Porto	UPORTO	Portugal
Felmica-Minerais Industriais S.A	FELMICA	Portugal
Ciaotech Srl	PNO	Italy

Besides the study area of this thesis (Tysfjord, Norway) the project has two more main study areas, Leinster in Southern Ireland and Wolfsber in Austria.

1.2. GIS (Geographic Information System) and Remote Sensing

Remote sensing (RS) and GIS have a long history of providing essential data to methodologically support environmental resource challenges (Pei *et al.*, 2021). This chapter discusses the main concepts and principles of remote sensing and GIS.

1.2.1. The GIS

GIS can be defined as an integrated set of Hardware and Software that can capture, organise, manipulate, analyse, model and present spatially referenced data that can be used to solve complex planning and management problems (Raper, 1991 *apud* Machado, 2000, p.240). GIS can also be defined as a set of tools that manipulates real world data for certain purposes (Burrough, 1986 *apud* Morais, 2013, p.17). Therefore, we can say that the GIS and its tools enable information to be generated and handled according to some certain objectives.

According to González *et al.* (2012) the GIS has five main components being the hardware, the software, the data, the methods and the human resources. To have a well-functioning GIS, these five components must work in harmony (Morais, 2013).

- (i) **Hardware**, which is the physical part of a computer encompassing parts such as the CPU (Central Processing Unit), storage units, input and output devices.
- (ii) **Software**, applications and operating systems that are appropriate for processing spatial information.
- (iii) **Spatial Data** whose main purpose is to aggregate and integrate georeferenced information from various sources such as remote sensing, cartographic surveys, aerial photography, censuses, among others.
- (iv) **Methods** which are the set of procedures, practices, and analyses that allow for more consistent and admissible results from a GIS;
- (v) **Human Resources** are an essential component to be able to deal with the complex procedures of this technology, from the treatment, survey, storage, presentation and analysis of data that require users with adequate skills and knowledge.

Also according to Martin, (1996) the operations that are inherent to GIS can be defined as:

- (i) Receipt, input and correction of data into the system which can occur through processes such as manual scanning rasterization, typing via keyboard and retrieval from other database systems.
- (ii) Storage and organisation that correspond to the control of the physical storage of data.
- (iii) Manipulation and analysis, which is the ability to manipulate spatial information and;

- (iv) Data output and graphical representation involving the export of data from the computer system into a human readable form.

The figure 1 shows us the operations of a GIS.

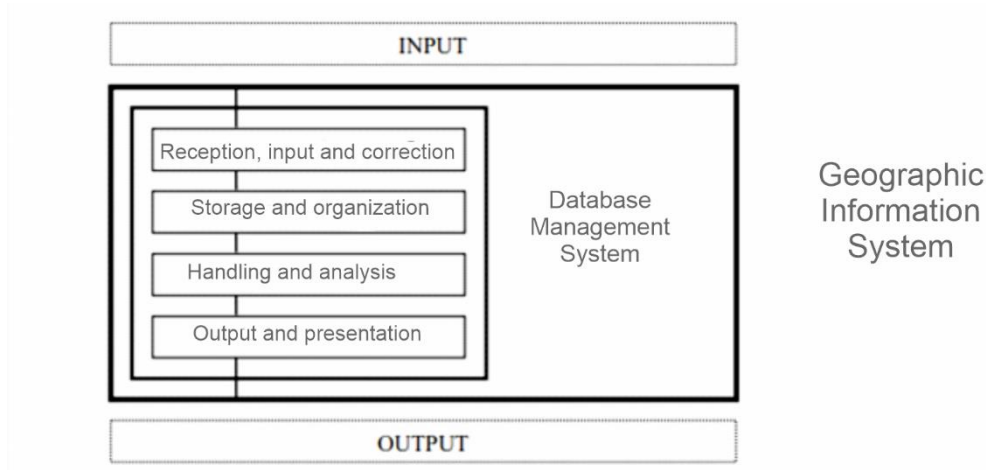


Figure 1. The GIS operations (Source: Martin (1996)).

1.2.2. Remote sensing

Remote sensing has several definitions among them we can mention Bakker *et al.* (2009), which defines RS as “the art, science, and technology of observing an object, scene, or phenomenon by instrument-based techniques”. Other authors such as Baghdadi and Zribi, (2017) define RS as “the acquisition of information on a given target without making direct contact with this target using the entire electromagnetic spectrum, from short wavelengths to long wavelengths”. The term 'Remote' is used because the observation is made without physical contact with the object of interest (Bakker *et al.*, 2009). Remote sensing has several sensors that vary according to the target to be observed (Toth and Józków, 2016). The remote sensor detects EM (Electromagnetic) energy, quantifies it and usually registers this energy in a digital format. According to Bakker *et al.* (2009), the energy or light source can be produced by the sensor itself (active sensor) like radar technologies or dependent on external lighting (passive sensors). Figure 2, gives examples of how these sensors operate.

To explain the process of obtaining information (figure 3), we can use as an example the interaction of the radiation emitted by the sun, which crosses the space interacting with the atmosphere and the surface of the planet, where is in part reflected back into space or absorbed

and, then, re-emitted in long wavelengths, which, in turn, are detected by the sensors as EMR - Electromagnetic Radiation (Souza, 2010).

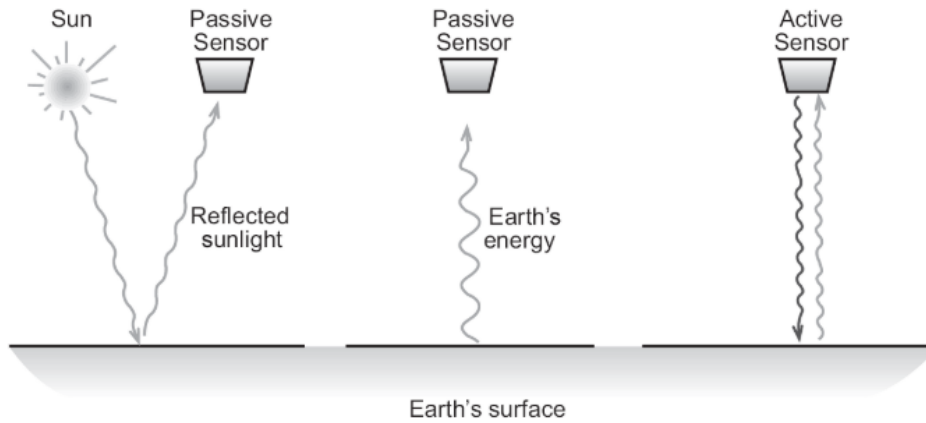


Figure 2. example of the functionality of passive and active sensors. (Source: Bakker *et al.* (2009)).

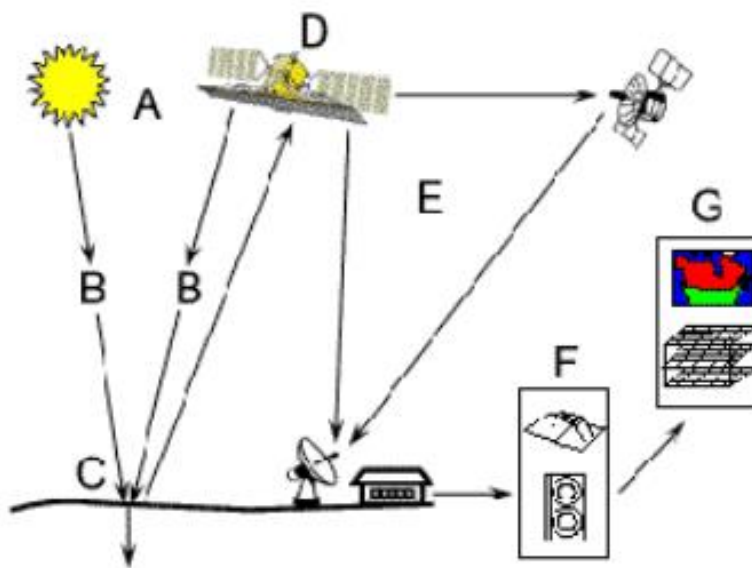


Figure 3. Exemple of the processes of EMR interaction with the Earth and obtaining information. (Source: Souza, (2010)).

The total range of wavelengths of EM radiation is called the electromagnetic spectrum. The electromagnetic spectrum is continuous and has no separate class boundaries, but can be separated into different portions as exemplified in figure 4 (Bakker *et al.*, 2009).

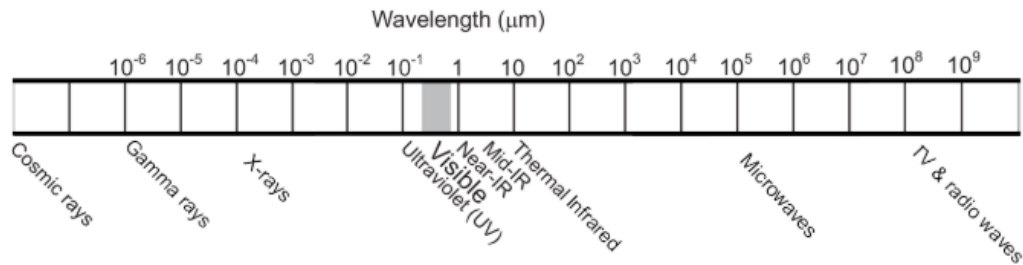


Figure 4. Representation of the electromagnetic spectrum in μm . (Source: Bakker *et al.* (2009)).

We can say that GIS tools are extremely important for remote sensing, allowing information from different sources, scales, accuracies and formats to be assimilated into a single source that can be used for modeling, mapping and decision support (Biswajeet and Saro, 2007; Jha and Tukkaraja, 2020). It is also important to note that the data manipulation, consultation, visualization and analysis attributes of GIS make these tools useful in generating maps of mineral favourability or potential (Harris *et al.*, 2000).

1.3. Pegmatites

After planetary accumulation and differentiation, the initial mineral evolution of the Earth's crust depended on a sequence of geochemical and petrological processes these processes, in turn, produced the first continents with their associated granitoids and pegmatites (Hazen *et al.*, 2008).

Pegmatites are defined by London (2021), as texturally complex igneous rocks marked by combinations of extremely large but variable crystal sizes and their intergrowth habits. Although pegmatites are associated with the most common plutonic igneous rocks, most of them, have granitic composition, being the most abundant pegmatites (London, 2021). Granitic pegmatites, in turn, are an important source of rare metals such as lithium, tantalum, niobium, beryllium, cesium, uranium, REE among others (Kesler *et al.*, 2012). The granitic pegmatites can be separated into three families (i) NYF family marked by the accumulation of Nb (niobium), Y (yttrium) and F (fluorine); (ii) the LCT family marked by the accumulation of Li (lithium), Cs

(caesium) and Ta (tantalum) and finally a mixed family (NYF + LCT)(Černý e Ercit, 2005). The pegmatites of Tysfjord, as it is described in item 1.4 of this thesis, are NYF pegmatites and rich in amazonite.

Although pegmatites occur worldwide, most gem-producing pegmatites such as topaz, tourmaline, aquamarine, morganite and kunzite are found in southeastern Brazil (more specifically in Minas Gerais) and also in Afghanistan, California (USA) and northwestern China (Oldershaw, 2005).

In the wide spectrum of structural patterns found in rare element pegmatites, we can distinguish them into three main types being: (i) homogeneous, (ii) zoned and (iii) layered (Cerny, 1991).

1.3.1. Homogeneous pegmatites

Homogeneous and coarse grained pegmatites with somewhat uniform distribution, with a few exceptions, these pegmatites mostly consist of elongated megacrysts of spodumene and quartz (Baldwin and Knorring, V., 1983; Cerny, 1991).

1.3.2. Zoned Pegmatites

For Cerny (1991), zoned pegmatites are the most diversified and common ones, that consist of different units with variable mineral modes. According to Heinrich (1948) we can distinguish the zones (figure 5) as (1) primary crystallization zones consisting of the intermediate zones (outer, middle, inner and margin of the Nucleo), and the core; (2) replacement bodies formed at the expense of pre-existing units under structural control; and (3) fracture fillings in dilated dislocations.

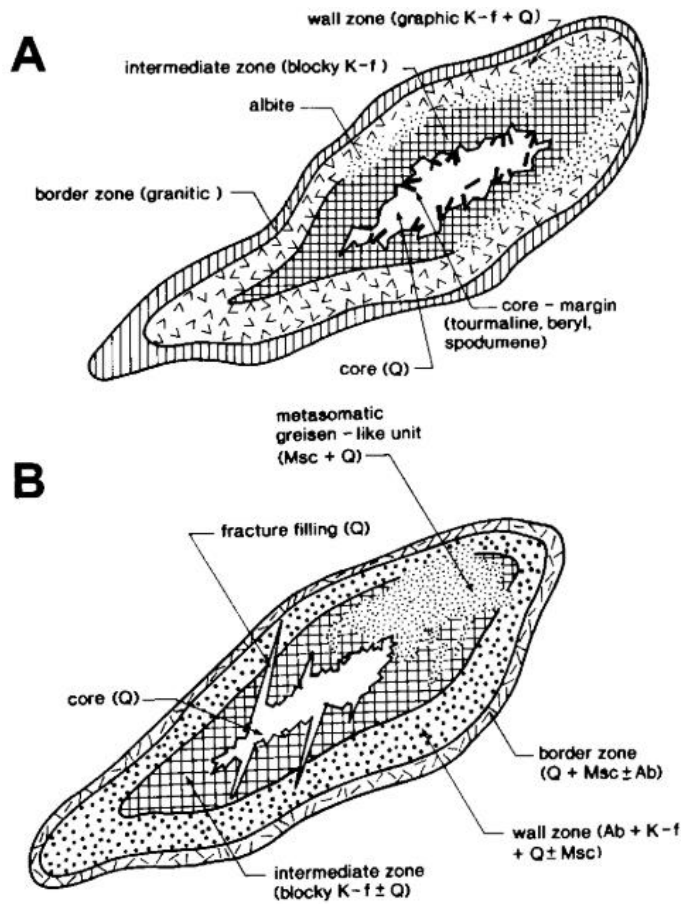


Figure 5. Internal structure of zoned pegmatites in horizontal sections (Source: Cerny, (1991)).

1.3.3. Layered pegmatites

Layered pegmatites consists of areas where minerals have been replaced. This replacement may be selective replacing individual minerals (such as topaz and micas), diffuse replacing primary minerals in inert zones, or maybe massive metasomatics replacing most of the primary assembly in pre-existing units (Cerny, 1991)..

1.4. Study Area

The study area is located in the Tysfjord district, an area of about 559 km² that is located in northern Norway. Due to the occurrence of many Nb, Y and F containing minerals, the Tysfjord-Hamarøy pegmatitic field, which includes the pegmatites of the study area, is classified as NYF family (Müller *et al.*, 2017; Hetherington *et al.*, 2021). First recognized in 1941 by Foslie (1941), these pegmatite fields are rich in amazonite, have large masses of fluorite and are intrinsically linked to the Tysfjord granites (Müller *et al.*, 2017). Of their 157 identified minerals, the main

ones are allanite-(Ce), fergusonite-(Y), columbite-(Fe), beryl, various sulfides and fluorite (Müller *et al.*, 2017). The Tysfjord-Hamarøy pegmatite field has several pegmatites identified, but those within the study area are namely: Jennyhaugen pegmatite, Nedre Øyvollen pegmatite and Håkonhals pegmatite (figure 6). The Jenny-Haugen and Håkonhals mines were chosen as the target of the first and second set of this study as they are large open-pit mines and are not limited by the spatial resolutions of satellites such as Sentinel-2 which have 10 m of spatial resolution.

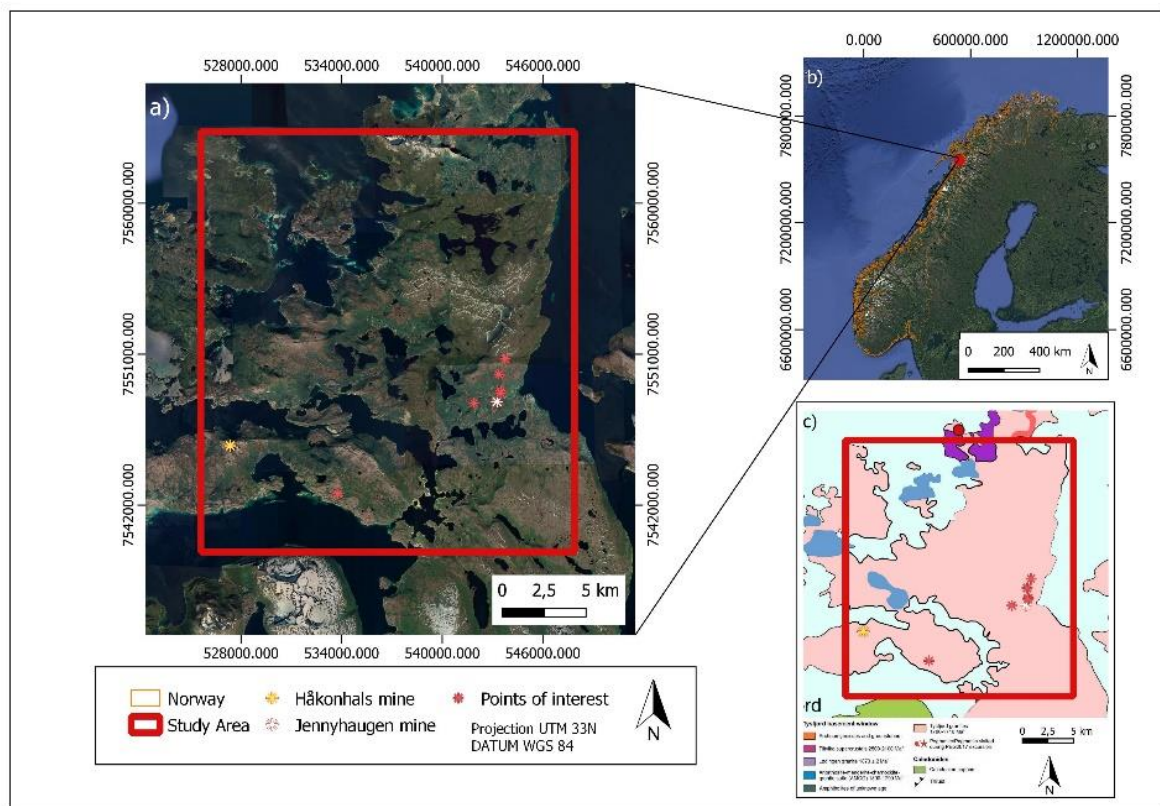


Figure 6. Location of the study area, Tysfjord, Norway. (Source: Modified from Müller *et al.* (2021)).

1.4.1. The Jennyhaugen Pegmatite

With an impressive NE-SW lens-shaped body and located at the coordinates: 543313 m E / 7548208 m N (WGS 84 - Universal Transverse Mercator (UTM) Zone 33N), the Jennyhaugen pegmatite has a width of up to 40 m and a length of at least 200 m (Müller *et al.*, 2017). Müller

et al. (2017) describes that when compared to other Tysjord pegmatites, the Jennyhaugen pegmatite is considered poor in minerals, this may be caused, according to the author, by the lack of available dump material. Of the materials found in old mines, quarries and dumps connected to this pegmatite, massive allanite-(Ce), monazite-(Ce), zircon, traces of yttrifluorite, arsenopyrite, löllingite with secondary pharmacosiderite, scorodite, native bismuth and 'amazonite' with inclusions of microlite, beryl and tantalite-(Mn) (Müller *et al.*, 2017).

1.4.2. The Nedre Øyvollen Pegmatite

Located at coordinates 543920 m E / 7548441 m N (WGS 84 - UTM Zone 33N), which corresponds to the active mine, the Nedre Øyvollen Pegmatite has two more locations that should be mentioned: the abandoned quarry (at coordinates 543483 m E / 7548669 m N) and the mine dumps (at coordinates 543346 m E / 7548397 m N) (Müller *et al.*, 2017). With about 30 x 60 m in horizontal section and approximately 80 m deep, this pegmatite has a vertical, cigar-shaped quartz core enclosed by a wall zone consisting of plagioclase, K-feldspar and mica. It also shows the following minerals: native bismuth, chalcopyrite, galenite, cassiterite, euxenite-(Y), calcite and yttrifluorite masses with variable contents of REE (Müller *et al.*, 2012; Husdal, 2008). According to Müller *et al.* (2012), this pegmatite has become a world-class and highly economical deposit, much of this thanks to its massive quartz body, the large crystal size and the low concentrations of trace elements.

1.4.3. The Håkonhals Pegmatite

Exposed at coordinates 527376 E / 7545507 N (WGS 84 - UTM Zone 33N), it is the pegmatite chosen as the target of study due to the large exposure of its open-pit mine (150 x 150 m). This mine is economically important due to the production of quartz with less than 50 ppm of contaminating elements, i.e., a highly pure quartz with very low contamination of other elements (Müller *et al.*, 2017; Hetherington *et al.*, 2021). The pegmatite is at least 25m thick vertically oriented horizontally, 400m long and 200m wide (Hetherington *et al.*, 2021). Müller *et al.* (2017), describes, at least, 28 minerals that can be found in the Håkonhals pegmatite among them we can mention quartz, bastnäsite-(Ce), allanite-(Ce), monazite-(Y), kainosite-(Y) and two types of yttrifluorite.

The Håkonhals and Jennyhaugen mines represented in figure 7, were chosen as a sampling target in this study due to the fact that they are large open-pit explorations/deposits making

them ideal for remote sensing identification, where there is always the limitation of the spatial resolution of the satellites used.



Figure 7. High Resolution Images of the Håkonhals and Jennyhaugen mines (Source: Adapted from Norge i Bilder (<https://www.norgeibilder.no/>)).

1.5. Structure of the work

The development of this thesis is organized in 4 chapters

Chapter 1. Introduction

In this chapter is introduced the theme where the main concepts of GIS and RS are specified and also the geographic framing of the study area.

Chapter 2. Metodology

In this chapter, is discuss the methodology used in this work. It specified the methods applied and how they were divided into three sets. The first corresponds to tradiconal methods already used for identification of LCT pegmatites. The second correpond to two traditional methods adapted for identification of NTF pegmatites in Tysfjord. The corresponds to the application of ML (Machine Learning) algorithms for the identification of pegmatites in the study area.

Chapter 3. Results and Disscusion

In this chapter the results obtained through the methods explained in the previous chapter are presented and analyzed in order to evaluate their effectiveness. The results are divided according to the three proposed sets and at the end a general evaluation of them is made.

Chapter 4. Results and Discussion

In this final chapter is presented a summary of the work and the main conclusions.

2. Metodology

In this work, it is intended, in addition to testing and evaluating the effectiveness of existing remote sensing methods, to develop new techniques specifically adapted to be applied in the Tysfjord district in order to identify NYF-type pegmatites.

Remote sensing has been applied to identify mineral deposits since the 1970s (Cardoso-Fernandes *et al.*, 2020^a). Some examples of its application are in the identification of uranium (Salles *et al.*, 2017) REE (Neave *et al.*, 2016; Zimmermann *et al.*, 2016), and especially, gold deposits (Amer *et al.*, 2016). Concerning pegmatite, we can cite the work of Cardoso-Fernandes *et al.* (2018^a), who since 2018 has been developing RS methods focused on studies of LCT-type pegmatites. Their studies have obtained important results for this field of research and for this work where we can highlight their results involving Sentinel-2 data (Cardoso-Fernandes *et al.*, 2018^a), new approaches in the detection of Li bearing pegmatite (Cardoso-Fernandes *et al.*, 2019^a), the development and application of ML algorithms in the identification of Li bearing pegmatites (Cardoso-Fernandes *et al.*, 2020^b) and the use of spectral data as a tool in remote exploration (Cardoso-Fernandes *et al.*, 2021). In addition to the studies by Cardoso-Fernandes, we can also cite the work of Gemusse *et al.*, (2019), who identified LCT pegmatites using image classification methods in Mozambique. Previous work has grounded the way for this research, their studies are focused on LCT type pegmatite and further focused mostly on lithium enriched LCT pegmatite.

The methodology employed in this study can be divided into three sets. The first consists in the application of traditional methods of image processing (RGB combination, Band ratios -BR and Principal Components Analysis -PCA) for the identification of LCT pegmatites. The second is the application of the traditional methods (in this case BR and PCA), but with the innovation that it is developed according to the mineral characteristics of the study area (a new approach). The third is the application of ML algorithms (SVM and RF). All these approaches were applied to optical satellite images.

In the first set, we apply the methods already used by Cardoso-Fernandes *et al.*, (2018^b), Cardoso-Fernandes *et al.*, (2019^b) and Santos *et al.*, (2019), such as image processing methods (BR, PCA and RGB combination) and ML algorithms (SVM and RF) in the identification of LCT pegmatites, aiming to evaluate the effectiveness of these methods in identifying NYF type pegmatites.

The second set consists of the development of two remote sensing methods adapted to the mineral characteristics of the Tysfjord district. To be able to adapt these methods, a spectral analysis was performed on samples obtained directly from the study area and after a spectral analysis, the most acquired bands from Sentinel-2 were selected and applied to BR, RGB combination and PCA. The workflow showed in figure 8, exemplifies the step-by-step process that takes place during the process.

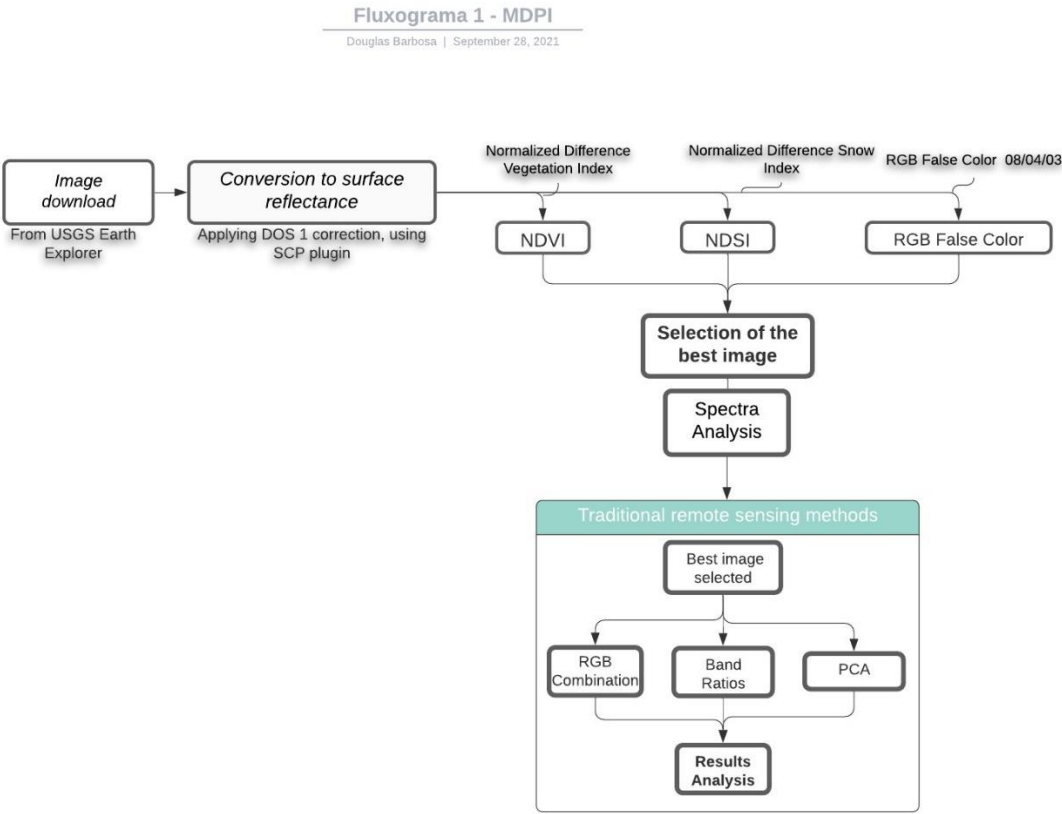


Figure 8. Step by step process for the development of a new approach for identification of NYF pegmatites.

The third and last approach, correspond to application of two advanced and robust ML algorithms when compared with the traditional image processing methods. The algorithms applied were RF and SVM, that is considered two of the best classifiers both with higher accuracy (Phan and Kappas, 2018).

2.1. Image Pre-processing

After downloading the images, some corrections such as atmospheric corrections, need to be applied. The atmospheric phenomena of scattering, absorption and refraction of the electromagnetic energy, end up affecting the radiance reflected by the surface, which in turn is captured by the sensor. Thus, atmospheric correction is necessary to minimize these atmospheric effects. Sentinel-2 images are already supplied in top of atmosphere reflectance. The atmospheric correction method applied was the Dark Object Subtraction called DOS1 (Chavez, 1996), which is a method that corrects the atmospheric scattering in which the atmospheric interference is estimated directly from DN (Digital Numbers) of the satellite image, was considered. The atmospheric correction was performed using the Semi-Automatic Classification Plugin (SCP) tool version 6.2.9 available in the QGIS software version 3.2.1.

2.2. Image Selection and Indices

All downloaded images have cloud cover less than 10% and after the atmospheric correction, the Normalized Vegetation Index (NDVI), Normalized Snow Index (NDSI) and RGB false-color composition were applied to select the images with less vegetation and ice cover. Images from the four seasons (winter, spring, summer and fall) from August 2015 to December 2020 were downloaded and compared. The selected image for this work was from 28-09-2019. This image was selected because it contains the least vegetation and snow cover compared to the other downloaded images.

2.3. Spectral Library and Analysis

2.3.1. Spectral Library

The spectral library was developed as part of one of the tasks of the Greenpeg project and its results were of great contribution to the spectral analysis of this thesis. Fifty-eight (58) spectra were collected from 21 samples from the study area. The samples were then dried in an oven at 48° C to remove any remaining moisture, which could affect the measurements. After drying the samples the spectra were collected using the ASD FieldSpec 4 equipment (figure 9). All spectra are acquired in UTF-8 text file format (.txt), a universal format accessible through several free softwares.

The nomenclature of the spectra is the same as the respective samples collected, when more than one spectrum is collected from a single sample, a number is added at the end that

corresponds to the order in which the spectrum was taken. For example, from the sample "T4220090125UIO" two spectra were taken, the first was named "T4220090125UIO_1" and the second "T4220090125UIO_2". The nomenclature of the samples also follows a standard stipulated by the Greenpeg project, where the first letter represents the area of study (T= Tysfjord, W= Wolfsberg e L= Leinster), the task in the ambit in which the sample was collected, the date in the year format in which the sample was collected and lastly the institution which collected the sample. After collection, the files were organized in a .xlsx file containing the name of the sample, the .txt file corresponding to the collected spectrum, the photo of the point where the spectrum was collected in the sample and the location of where the sample was collected and the coordinates of the collection point.



Figure 9. Spectroradiometer used to collect the spectral data of the Tysfjord samples.

2.3.2. Spectral Analyses

For Rayner (2001), spectral analysis is the analysis of the oscillation or calculation of waves in a set of sequenced data. Spectral analysis has been used for several purposes in geology, such as studies involving gold mining (Amer *et al.*, 2016), minerals in general (Panteleev *et al.*, 1995; Patel *et al.*, 2020), RRE (Martin, 2015) and more recently Cardoso-Fernandes *et al.*, (2021), conducted a study focused on the spectral analysis of Li minerals.

In this work, fifty-eight spectra were analyzed, of which five (5) correspond to amazonite, thirteen (13) to potassium feldspar, nine (9) to plagioclase, sixteen (16) to quartz core, six (6) to granite, four (4) to biotite and five (5) have no defined minerals. The analysis was done by cross-referencing the absorption and reflectance information of the samples with the spectral range that the Sentinel-2 satellite bands cover.

The analysis was done by cross-referencing the absorption and reflectance information of the samples along the wavelength with the spectral range that the Sentinel-2 satellite bands cover at the same wavelength. The figure 10 gives an example of how this analysis was performed. Figure 11 shows the spectral coverage of the satellite bands used for this analysis.

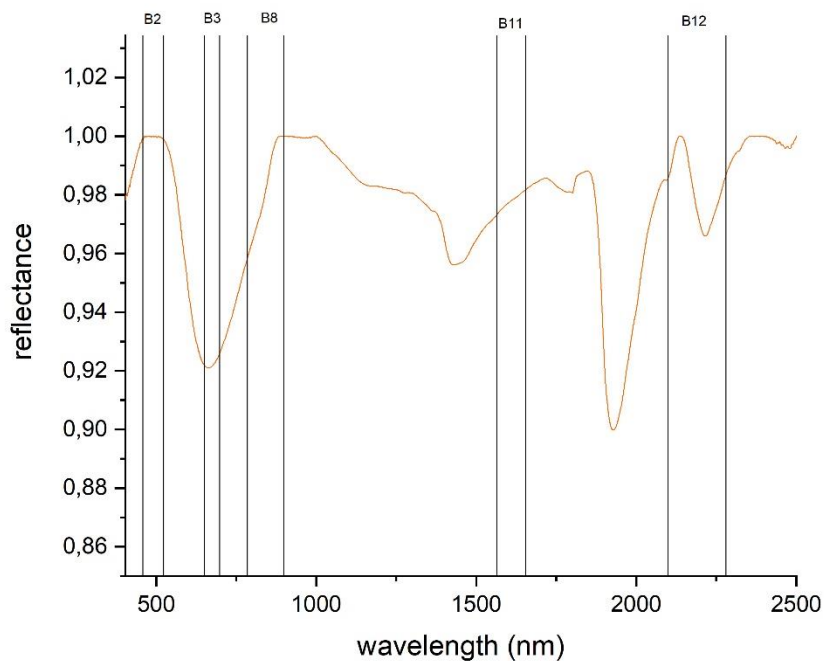


Figure 10. Spectrum of a sample of amazonite from Tysfjord in relation to the sentinel 2 satellite bands.

Band	Sentinel-2A MSI		
	Spectral region	Wavelength range (nm)	Resolution (m)
B1			
B2	Blue	458–523	10
B3	Green peak	543–578	10
B4	Red	650–680	10
B5	Red edge	698–713	20
B6	Red edge	733–748	20
B7	Red edge	773–793	20
B8	NIR	785–899	10
B8A	NIR narrow	855–875	20
B11	SWIR	1565–1655	20
B12	SWIR	2100–2280	20

Figure 11. Wavelength for sentinel 2 (Source: adapted from Cerasoli *et al.*, (2018)).

2.4. Image Processing

2.4.1. RGB combination

In order to produce different color combinations, different bands can be combined and used for different objectives (Pour and Hashim, 2015; Cardoso-Fernandes, *et al.*, 2019^a). The RGB combination tested in the first set (RGB 3-2-12), was developed and applied by Cardoso-Fernandes *et al.*, (2019^a), for the identification of pegmatite enriched with Lithium minerals using the Sentinel-2 satellite bands. In the second set, the RGB combinations were generated through the insertion of the BRs into the red, green and blue channels, in an attempt to form new combinations that highlight the NYF pegmatite of the study area. The combinations tested in the second set were 4/8, 2/8 and 8/12, 4/8, 3/8 and 8/12 and 4/8, 8/4 and 8/12. Following the method applied for Cardoso-Fernandes, *et al.*, (2019^a), the order of the BR introduced is according to the objective target that should be highlighted. In the red channel, the band or BR that most reflects the target is introduced, in the blue channel, the one that absorbs the spectra of the target or reflect granite areas and in the green channel the best combination is sought to obtain an acceptable image to be analyzed. In the red channel was introduced the BR that reflected the known pegmatite areas.

2.4.2. Band Ratios

One of the most widely used image processing methods for lithological purposes, BR, consists of splitting two bands to highlight specific spectral differences (Simon et al., 2016). In several studies (Goetz, 2016; Cardoso-Fernandes, 2018^b; Cardoso-Fernandes *et al.*, 2019^a; Santos *et al.*, 2019), the division between two bands was performed and the result analyzed individually. For the first test, the band ratios applied is the ratio 3/8 developed for Spodumene and 4/12 for Lepidolite, considering Sentinel-2 data.

Now for the second set, after the spectral analysis and selection of the main bands of interest for this study, a series of band combinations were tested and the best ones were selected. The BR tested are shown in table 2.

Table 2. Band Ratios tested for the second set, considering Sentinel-2 data.

Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	Band 8	Band 11
2/4	3/4	4/3	5/3	6/3	7/3	8/3	11/3
2/8	3/8	4/12	5/12	6/12	7/12	8/4	11/12
2/12	3/12		5/4	6/4	7/6	8/6	
2/3	3/2				7/4	8/12	

2.4.3. Principal Components Analyses (PCA)

Principal components as a multivariate statistical technique used to enhance and separate certain types of spectral signatures from the background (Singh and Harrison, 1985; Loughlint, 1991). This method was created by Hotelling in 1933, to be used as a support in your work in the field of educational psychology (Singh and Harrison, 1985). Over the years, it has been refined and applied in other areas and is now used, for example, for image enhancement, for digital change detection and for determining the underlying statistical dimension of the image data set (Byrne *et al.*, 1980; Gonzalez and Wintz., 1977; Singh and Harrison, 1985). The formula that expresses the principal components (PC) is:

$$Y_j = a_1 X_1 + a_2 X_2 + \dots + a_N X_N = a_j^T X \quad (1)$$

where T denotes transpose of a matrix and $a_j = [a_{1j}, \dots, a_{Nj}]$ are the normalized eigenvectors [i.e., $a_j^T a_j = I$] of the variance-covariance matrix (Singh e Harrison, 1985).

The PCs tested were all two-band components, because they were the ones that had obtained the best results in previous work. The PC tested for the first set is the PC of bands 3,8. For the second set, the PC's tested were from bands 2,8; 3,12; 2,12; and 8,12.

2.5. Machine Learning (ML) Algorithms

Two algorithms ML applied were SVM and RF. These algorithms were optimized and evaluated in Python programming language using a free open-source library for ML called scikit-learn (Pedregosa *et al.*, 2011). As it can be observed in figure 12, step by step, the processes were created in a way that its possible to minimize errors and always improve the algorithm, because of this the processes have two "stops".

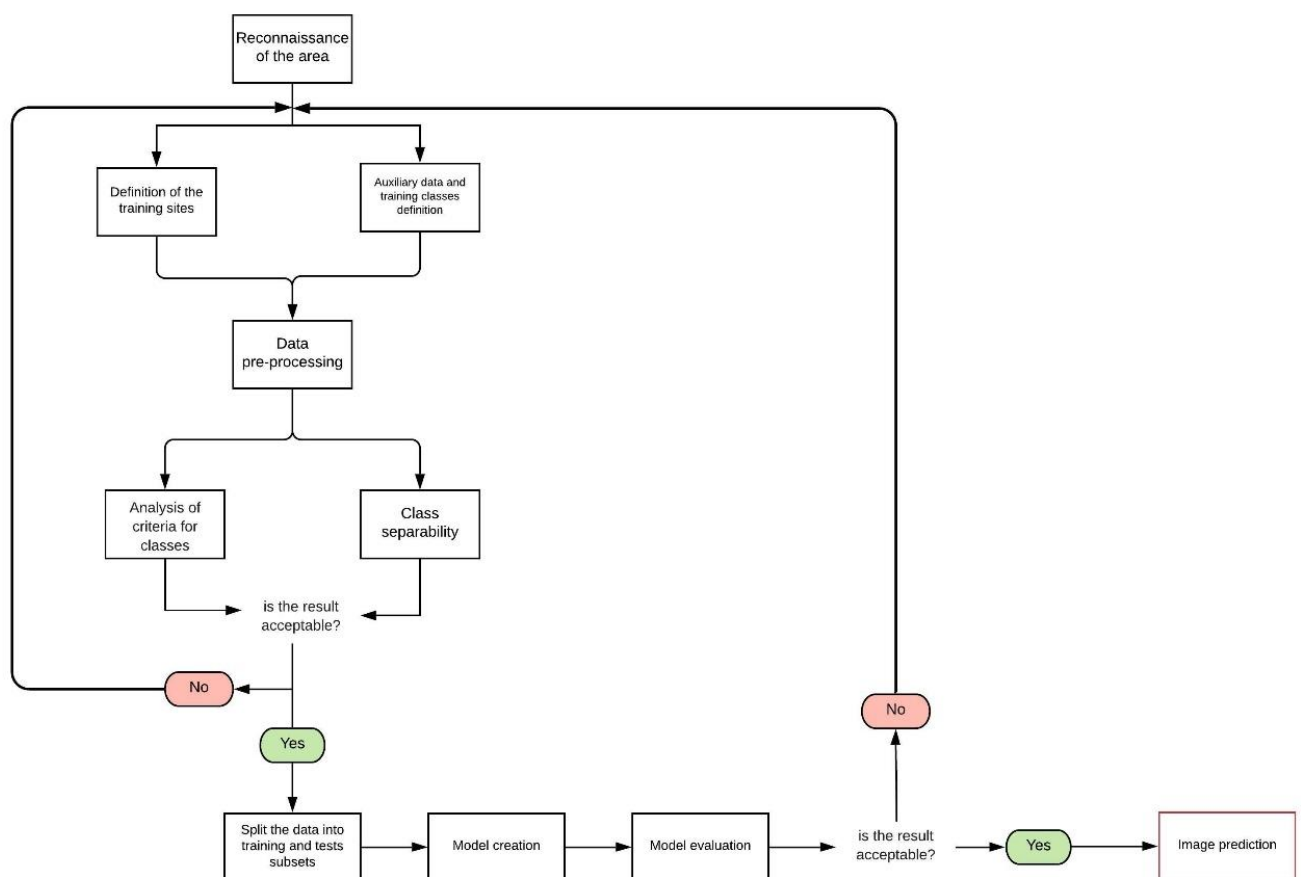


Figure 12. Workflow of the ML application procedure.

The first "stop" occurs after the Analysis of criteria for classes and Class separability. In this stop, two parameters are taken into consideration. The first is the class separability, where is check if the classes are well separated and distinct, thus preventing pixels from being misclassified. The

class separability was analyzed in the software PCI Geomatics (2018). If one or more classes presented low separability, the classification is remade. The second parameter consists of verifying that the training samples meet the standards specified in the literature for a good classification. Some authors such as Eastman JR (1998), suggest that each training set should be no less than ten times the number of bands. Chen and Stow (2002) highlight the importance of distributing the training samples as evenly as possible in spatial terms to obtain good results. When the parameters described above are not acceptable, the training samples are resampled and the process is repeated until the desired result is obtained.

The second “stop” comes after the model evaluation phase, because in the previously employed two-stage grid-search the quality of the algorithm training was evaluated through a cross-validation process. If the cross-validation result is not as expected, we go back, redo the training samples and repeat the whole process until the best result is obtained.

Several models were created with different variables such as: (i) class imbalance, (ii) class balanced, (iii) data splitting considering the regions of interest (ROIs) as proposed by Cardoso-Fernandes *et al.* (2020^b) and; (iv) data splitting without considering the ROIs. Of these models, the one that presented the least overfitting was chosen.

2.5.1. The Random Forest Algorithm

A very robust algorithm for the classification of remote sensing data from multiple sources, RF is a supervised classification method that is much lighter than boosting-based methods and lighter than conventional bagging with a comparable tree-type classifier, capable of improving classification accuracy (Gislason *et al.*, 2006). 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 (Pavlov, 2020). To ensure the randomness of the trees in the ensemble, a different subset of multispectral bands can be select to use when splitting each node during the tree construction (Müller, 2016). Taking this into account, two main parameters need to be optimized: (i) the number of trees in the ensemble and; (ii) the number of features used in each node for splitting (Phan and Kappas, 2018).

2.5.2. The Support Vector Machine Algorithm

SVM is an ML algorithm developed by Vapnik (2000) and based on statistical learning theory (Yu *et al.*, 2012). Being one of the most popular ML being able to provide linear or nonlinear

classification, regression, and even outlier detection being considered, therefore, a very powerful and versatile ML model (Geron, 2019). For Cardoso-fernandes *et al.* (2020^a), the SVM technique separates the dataset into a defined number of classes by trying to find a supporting hyperplane which is a decision boundary that maximizes the distance of the edge between the class boundary hyperplanes. Phan and Kappas (2018), concluded in their work that the SVM classifier produces, on average, the largest overall accuracy and has less sensitivity to the size of the formation samples. As pegmatites typically have small bodies, which consequently generates small formation samples, SVMs become a powerful resource for classifying these areas.

2.5.3. Reconnaissance of the area

Auxiliary data and definition of the training classes

The GIS data used in this study was, in part, provided by the Greenpeg project partners. These data were analysed and the more suitable data were selected for use. In the .shp file of pegmatites location, for example, only five 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. Beyond the GIS data, the selected data also includes geological maps (Müller *et al.*, 2021), aeromagnetic data (Olesen *et al.*, 2010), high-resolution aerial photographs (NIBIO), the best of traditional methods (RGB combination, Band Ratios and PCA). NDVI were used to understand, identify and to define the classes presents in the study area.

At first, six classes have been defined for the Tysfjord case study:

1. Pegmatites bodies,
2. Caledonian nappes,
3. Granite,
4. Ice/Snow,
5. Water Bodies and;
6. Vegetation,

but due to problems with overfitting and false positives in the algorithms, the study area was reduced from Tysfjord province to Tysfjord district.

After that, the classes were reduced to four classes being:

1. Pegmatites bodies,
2. Granite,
3. Water Bodies and;
4. Vegetation.

The sample locations (or training sites) were selected considering the following criteria (i) quality of exposure; (ii) pixel purity, and; (iii) a uniform distribution, as possible, over the image. In this study, we used five known pegmatite areas as target samples. Of these five, two are namely Håkonhals and Jennyhaugen mines and the other three are located in the area of the Nedre Øyvollen Pegmatite.

Some authors (Fitz *et al.*, 2019; Richards and Jia, 2006), suggested that each class have between 30 and 60 samples and Eastman JR (1998), argued that the number of pixels each training set should not be less than ten times the number of bands. Respecting the parameters suggested, the four classes defined have at least 30 sampling polygons and the number of pixels for each class is more than 10 times de the number of bands (five bands). Table 3, presents the relationship between the number of pixels and samples per class.

Table 3. Relationship between the number of pixels and samples per class

ID	Class	poly x class	pixel x class
1	pegmatites bodies	30	79
2	granite	30	356
3	water bodies	31	181
4	vegetation	39	182

The spectral analyses made is possible to identify five bands more suitable for geological proposes. That bands are bands 2, 3, 4, 8 and 12.

2.5.4. Class Separability

In order to confirm if the classes are well separated, avoiding pixels from being misclassified, the class separability was analyzed using the PCI Geomatics software. The class identification is redone if one or more classes present low separability. Tables 4 and 5 show the class separability for the first training samples.

Table 4. 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 5. 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 was made without applying the parameters suggested previously. Tables 6 and 7 show the class separability after employ the parameters suggested, where it is possible to see that the less separability corresponds to pegmatites and granite classes.

Table 6. 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 7. 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	

Is clear, as it is possible to see in Tables 6 and 7, that the separability result is better after following the recommended parameters. The smaller class separability, was between pegmatites and granite classes (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 4 and 5). All the other classes presented maximum separability.

2.5.5. Split the data into training and tests subsets

The dataset was split, with 25% of the samples split into testing, and the remaining 75% split for training. The data split was made according to their respective region of interest as proposed by Cardoso-Fernandes *et al.* (2020^a), to ensure the maximum independence between the training and test subsets.

2.5.6. Model creation

The first step in the model creation was to define the parameters to optimize and its correspondent parameter range (or variation) to be used in an automatic search (grid-search). For the RF in a preliminary stage, only the number of trees, defined by the function 'n_estimators', was tested. In this work, 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 Tysjford study case. For Sentinel-2 data with the RF algorithm, the optimal number of trees was 50. Figure 13 and 14 shows the cross-validation score (accuracy in this case) for the number of trees tested.

```
Fitting 5 folds for each of 11 candidates, totalling 55 fits  
Best parameters: {'n_estimators': 50}  
Best cross-validation score: 0.97  
Test-set score: 0.94
```

Figure 13. Best parameters achieved for the RF algorithm, using Sentinel-2.

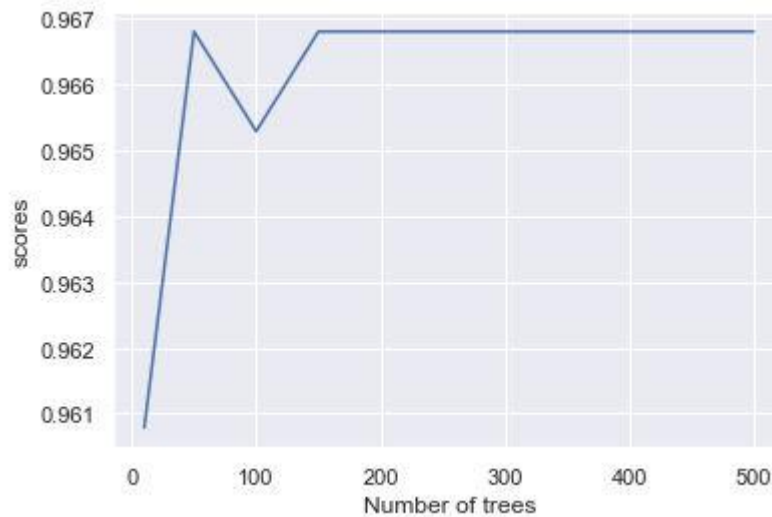


Figure 14. Cross-validation score as a function of the tested parameter for Sentinel-2.

In the case of the SVM algorithm, for the same study area and using the same data, figure 15 shows that the best result in the grid search is located in line 1 (for the C parameter) and line 100 (for the gamma parameter). The grid search represented in figure 16 shows us that for the best parameters the cross-validation score achieved was 0.97.

```
Fitting 5 folds for each of 36 candidates, totalling 180 fits
Best C: 1
Best gamma: 100
Best cross-validation score: 0.97
```

Figure 15. Best parameters location in a grid search for the SVM algorithm.

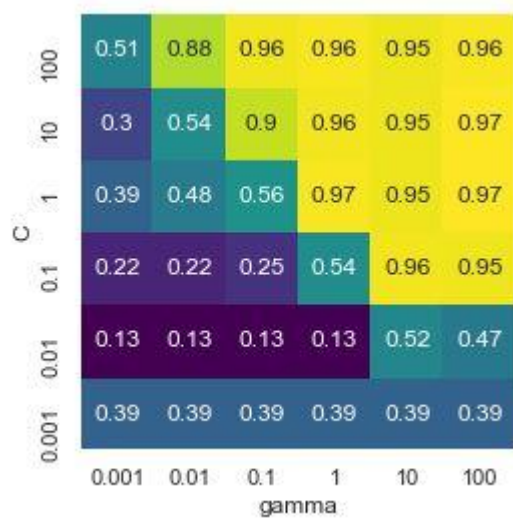


Figure 16. Cross-validation score obtained for each parameter combination.

2.5.7. Model evaluation

At this stage, the best model automatically returned in the previous step, is evaluated using the test subset. Different evaluation metrics can be employed, including the overall accuracy (OA) and/or kappa statistics. Having evaluated the model performance on the test set and achieving acceptable results, it is possible to move on to the classification process and predict the whole image using the model created.

3. Results and Discussion

3.1. First set: Results for LCT traditional methods

3.1.1. RGB Combination

The RGB combination tested was the combination 3-2-12 for Sentinel-2 image. As we can see in figure 17, this RGB combination highlights the pixels with NYF pegmatites in white color. Also, as we can see in figure 17 b), a signal confusion is observed with the granites bodies at Håkonhals mines which was also classified in white pixels. The Jennyhaugen mine (figure 17 c)) in its turn was well highlights in white pixels without signal confusion with granite bodies. This may be possible due there are no granite outcrops around this mine, which is surrounded by vegetation. Other elements of the study area are highlighted like vegetation in purple color, water bodies in dark pixels, and ice/snow in yellow pixels.

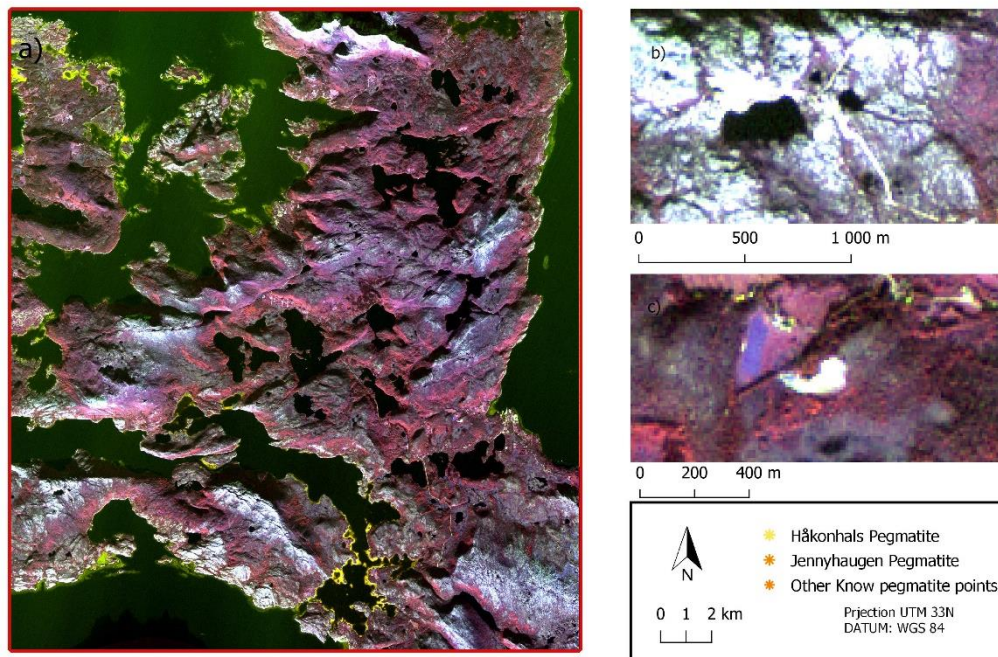


Figure 17. RGB composition of bands 3-2-12 for Sentinel-2.

When applied to the identification of LCT pegmatites (Cardoso-Fernandes *et al.*, 2018^b; Santos *et al.*, 2019), this RGB composite gives a result different from what could be observed in Tysfjord.

In the region of São João del Rei, in the state of Minas Gerais, Southeast Brazil, granite is highlighted in dark purple and water in greenish tones (figure 18), while in Spain, granites appear in blue or greyish blue. In both cases known pegmatite bodies are highlighted in yellow, but in Brazil, as is showed in Figur18, some areas of known pegmatite bodies are highlighted in white pixels (figure 18 c) and d)).

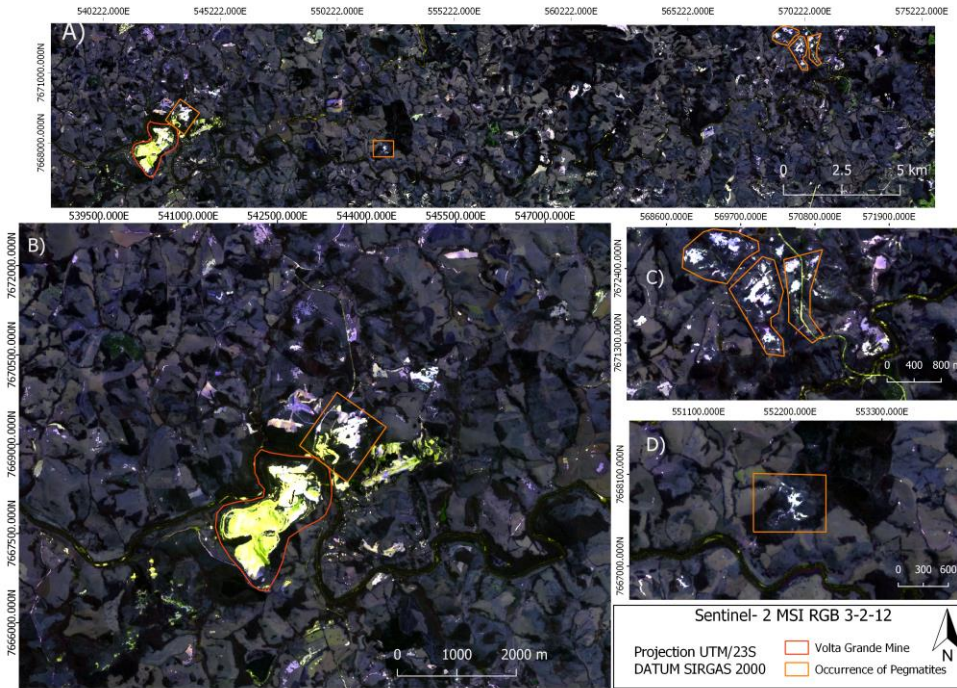


Figure 18. RGB Combination 3-2-12 for Sentinel-2. (Source: Santos *et al.* (2019)).

3.1.2. Band Ratios

The best BR for this set were the ratios with the bands 3/8. This ratio highlights the two principal know pegmatite points (Håkonhals and Jennyhaugen mines) in light grey pixels. Granites bodies and vegetation appear in dark pixels and water bodies in light pixels (figure 19). This ratio performed best in Brazil, where it highlighted areas with the presence of LCT pegmatite in brighter pixels (figure 21).

The BR 4/12, in turn, fails to highlight the known pegmatite areas. In both Håkonhals and Jennyhaugen mines was a strong noise making it difficult to analyze (figure 20). In both cases, there was a signal confusion with the water. In the identification of LCT pegmatites, in Brazil, this ratio presents a completely different result. Even if it has not highlighted pixels as well as the ratio 3/8, it has highlighted them in a grey color (figure 22).

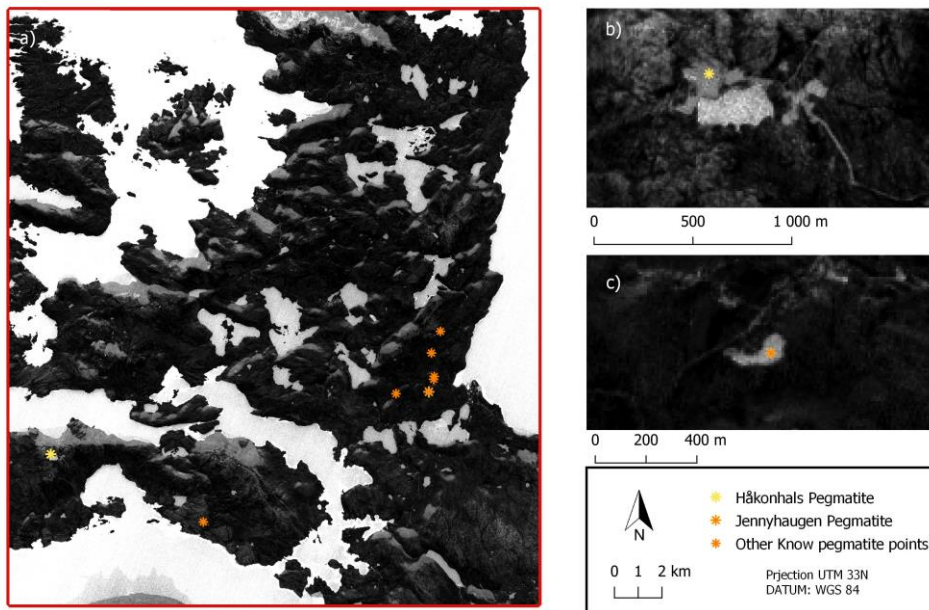


Figure 19. Band Ratio 3/8 for Sentinel-2.

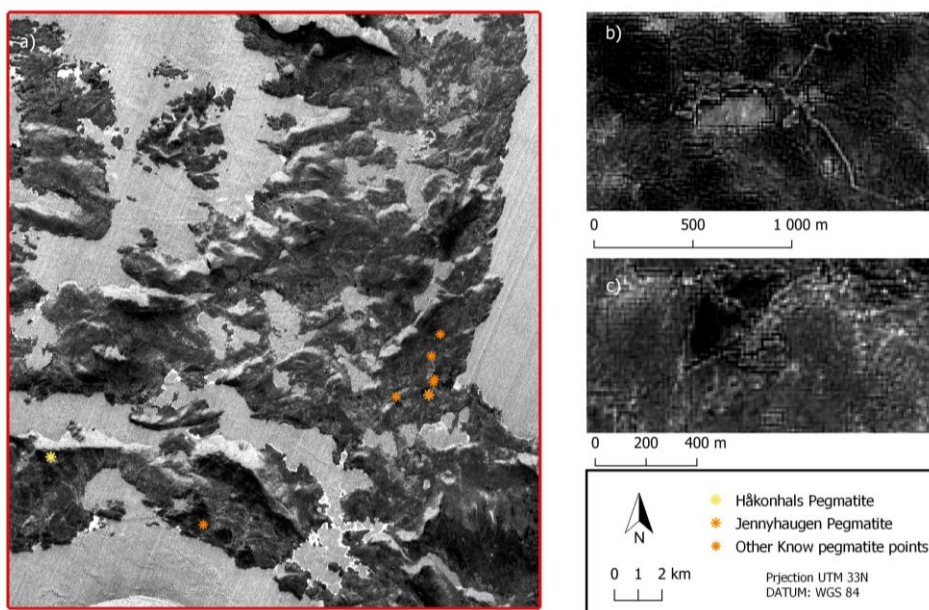


Figure 20. band Ratios 4/12 for Sentinel-2.

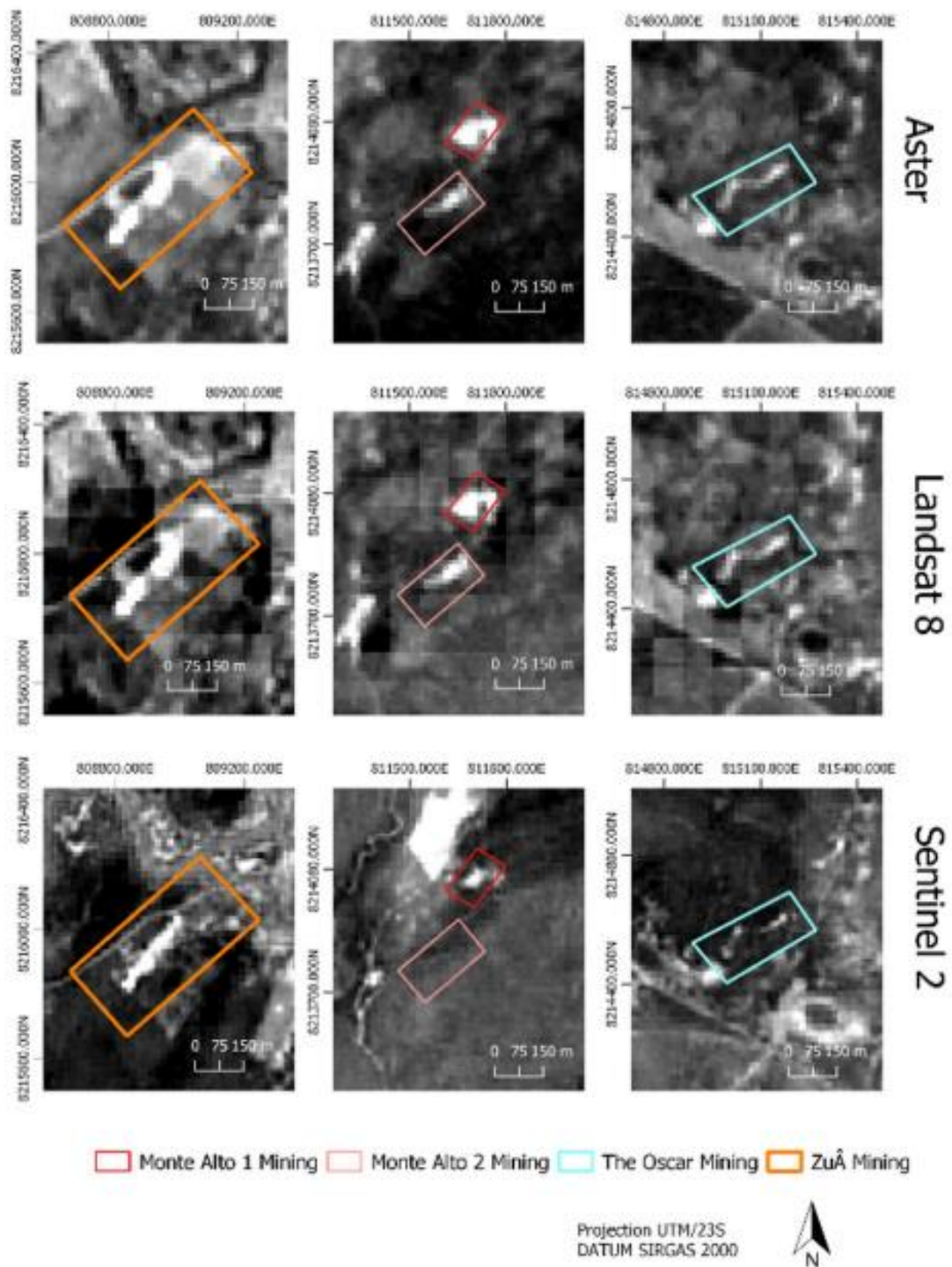


Figure 21. LCT pegmatite highlighted in Brasil. BR 3/8 for Sentinel-2 bands. (Source: Santos *et al.* (2019)).

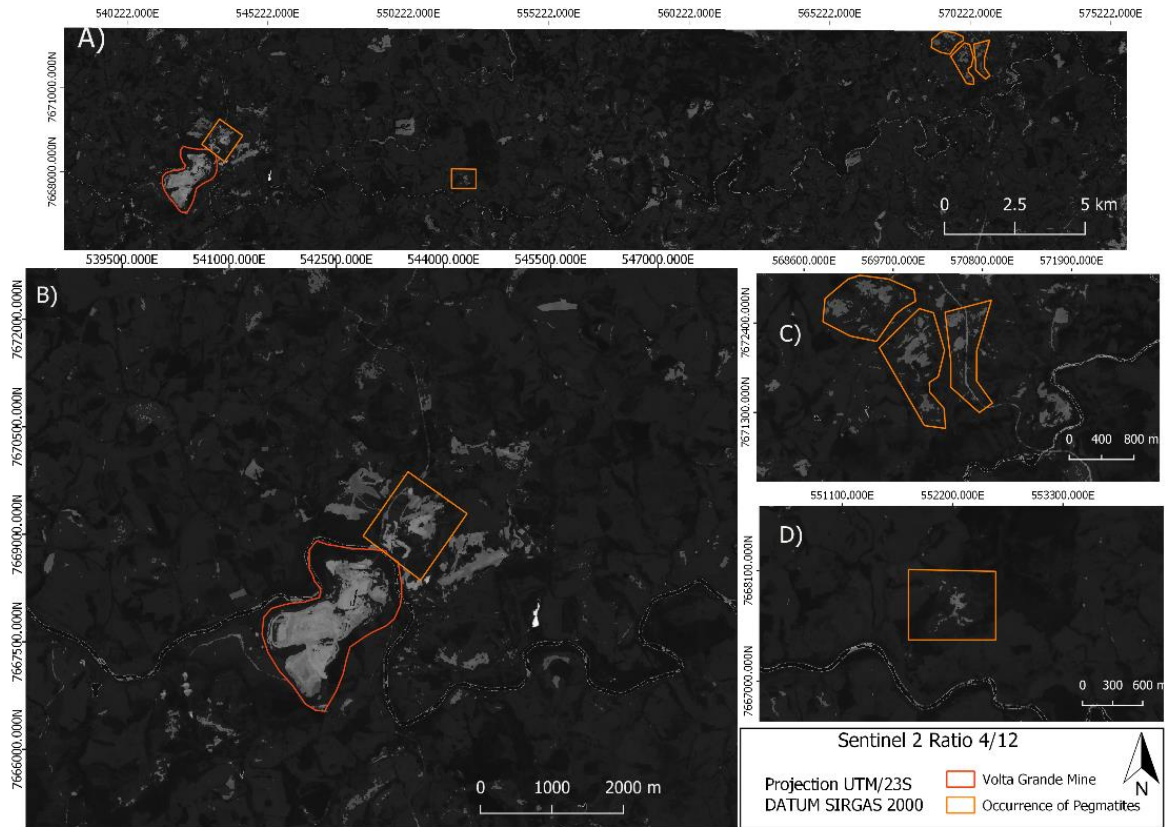


Figure 22. LCT pegmatite highlighted in Brasil. BR 4/12 for sentinel 2 bands. (Source: Santos *et al.* (2019)).

3.1.3. Princpail Components Analyses (PCA)

The PC two bands tested (bands 3 and 8) had an excellent result in identified NYF pegmatites bodies. As it is possible to see in figure 23, both mines (Håkonhals and Jennyhaugen), were very well highlighted and were very well distinguished from the surrounding elements with no signal confusion. In the PC2 presented in figure 23, other elements of the study area were highlighted in ranging from light grey to black pixels. The vegetation is highlighted in tones of grey pixels. The water bodies in light grey pixels and granite bodies in grey color. One detail that could be an obstacle in this method is the classification of ice/snow, which is highlighted in light pixels which can cause signal confusion in an area with more presence of snow. This selective PC did not perform as well in Brazil (figure 24), where they could not highlight the known pegmatite areas as well, which were highlighted in dark gray.

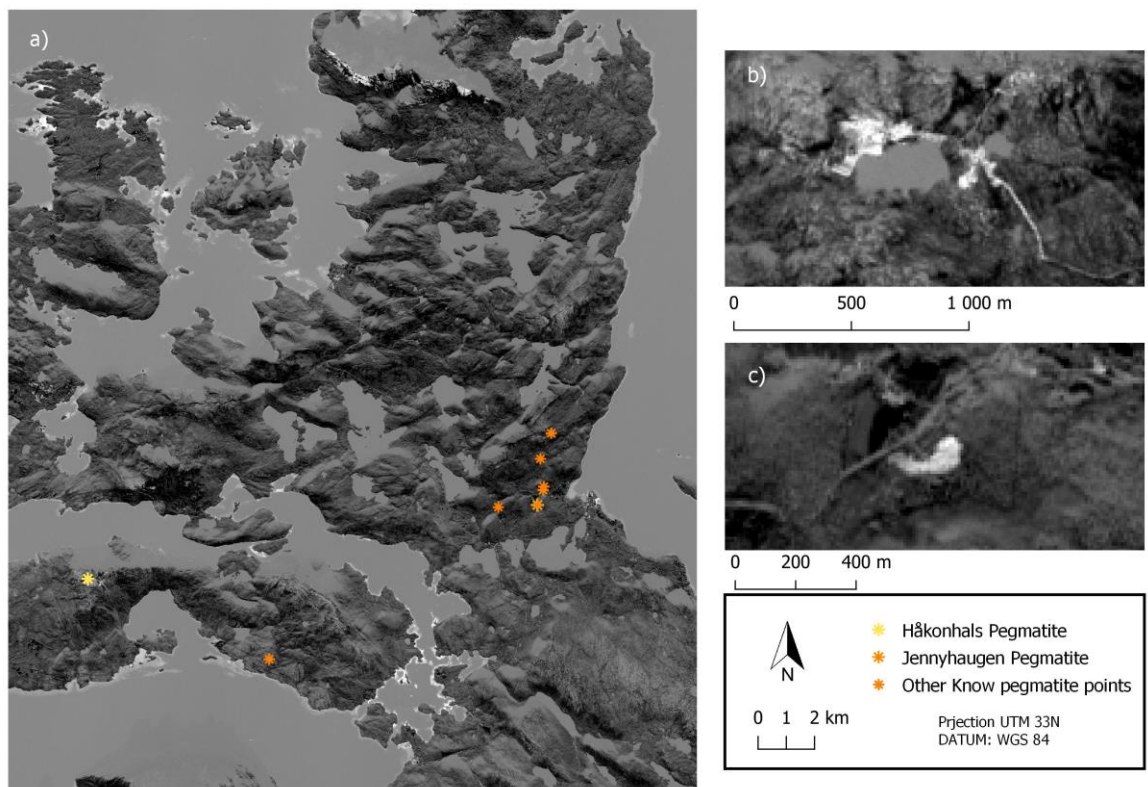


Figure 23. PC 2 bands of the bands 3,8 for Sentinel-2.

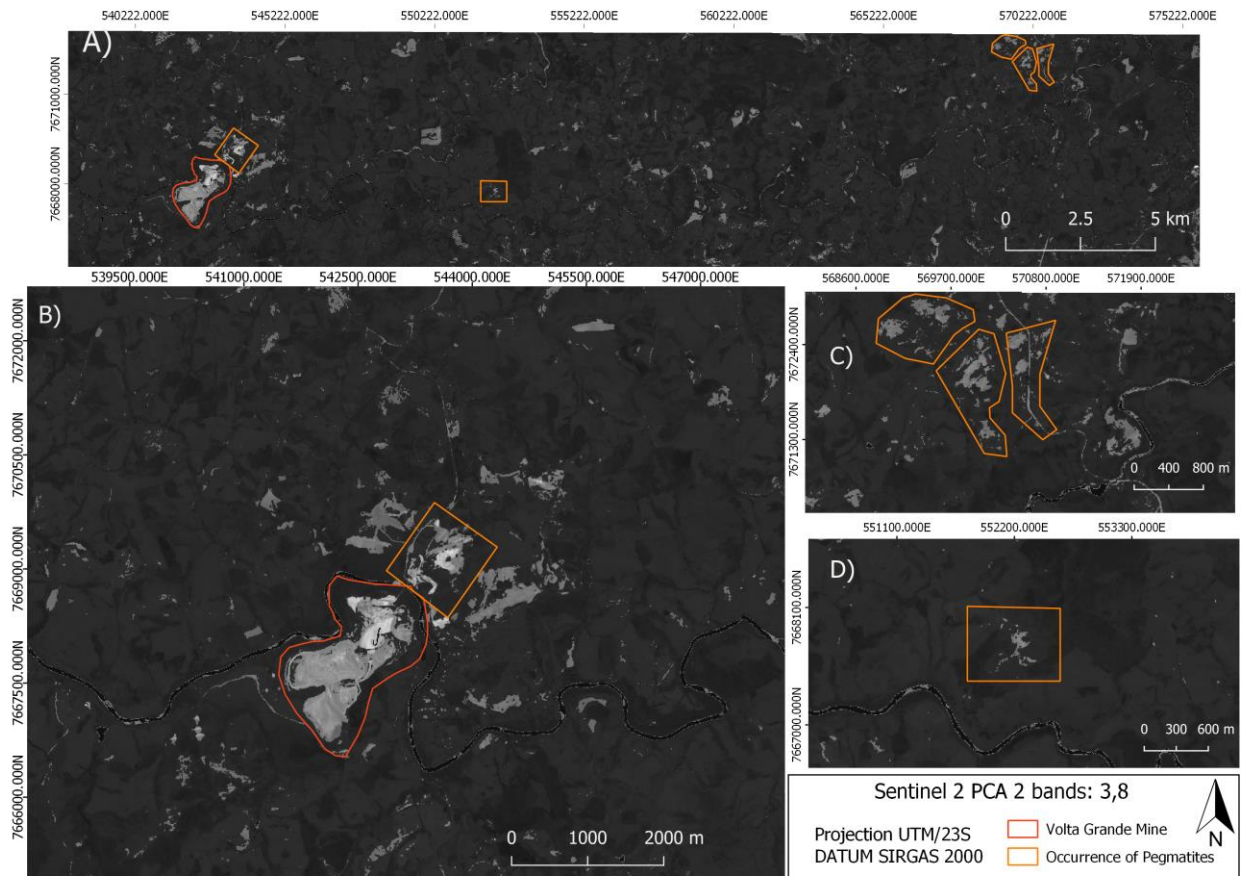


Figure 24. Selective PCA (2 PCs) for Sentinel-2. (Source: Santos *et al.* (2019)).

3.1. Second set: A new approach for traditional methods.

In this section will be presented a new approach for traditional methods adapted for the characteristics of the study area. The methods adopted were PCA and BR, this last one in turn was applied in the RGB channel to generate new RGB composites. This set is focused on two areas of know pegmatites (Håkonhals and Jennyhaugen mines). These points were chosen being into account that these mines have a very quality of exposure being very large mines and being less affected by the spatial resolution of the Sentinel-2 satellite.

3.1.1. Principal Components Analyses

The selective PC that showed the best result was the PC of bands 2,8. This PC managed to highlight well the mines area, without signal interference from other map elements like water and granite. As we can see in the figure 25, this PC highlights the pixels of the mines in white, granite in grey, while water and vegetation are highlighted in dark pixels.

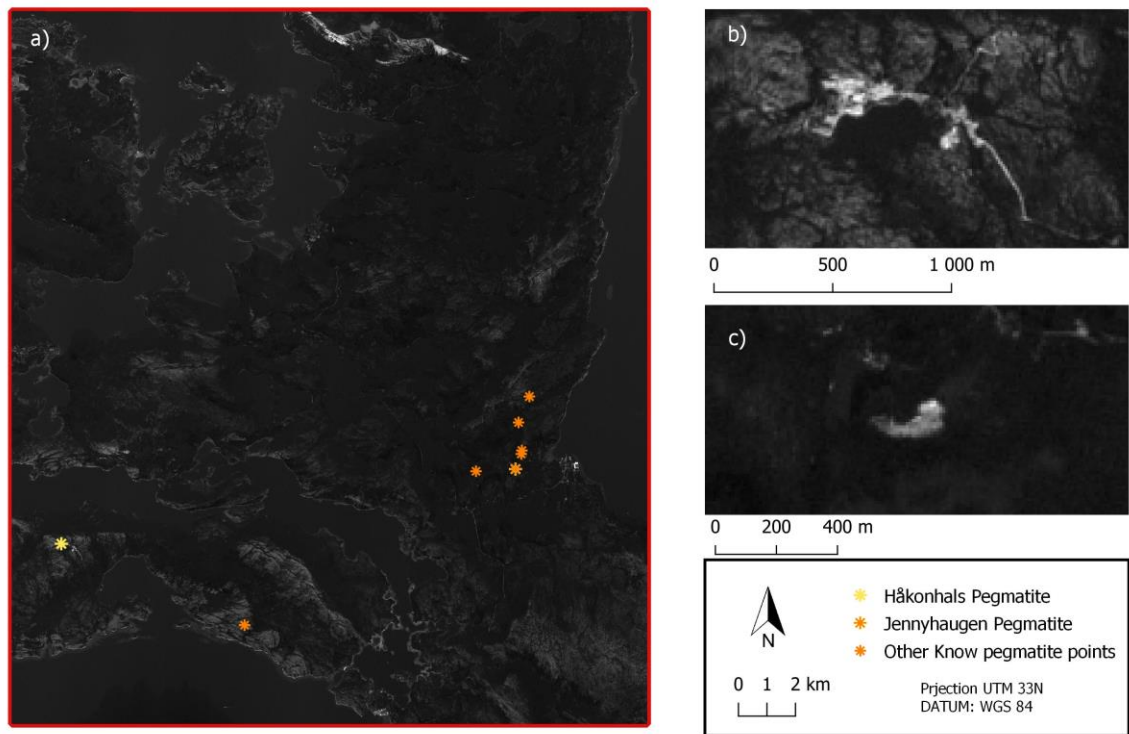


Figure 25. PC 2 bands of the bands 2,8 for sentinel 2.

The other PC's, namely the PC's of bands 3,12 and 2,12 obtained a similar result identifying the pegmatites pixels in light colors and the remaining elements in greyish or dark pixels. Thus, they were able to differentiate the pegmatites well from the other elements around them (figure 26). The PC 8,12 (figure 27) did not obtain acceptable results, with strong noise and not being able to identify the areas with known pegmatites.

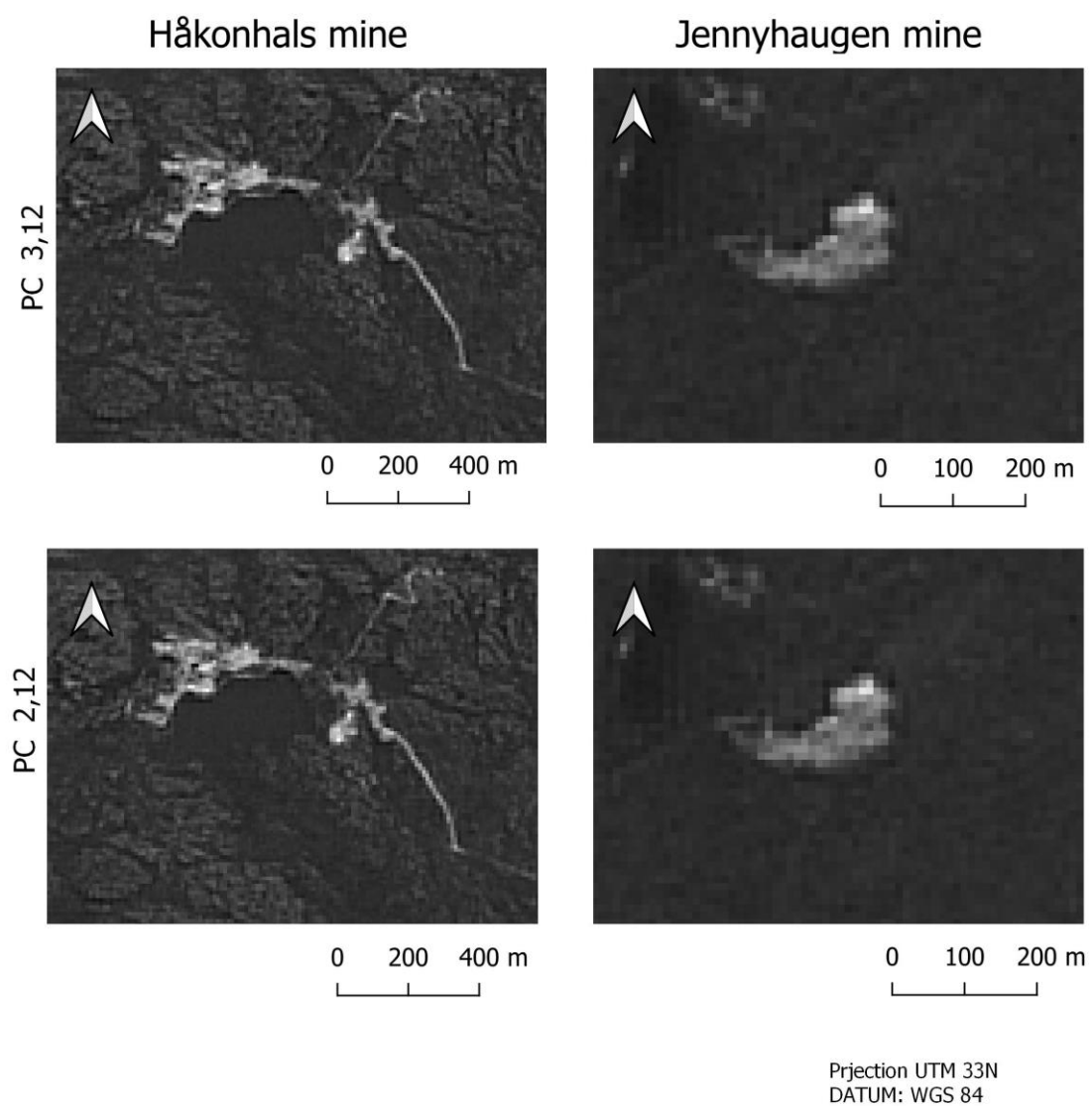


Figure 26. Comparison of the result between PC 3,12 and 2,12

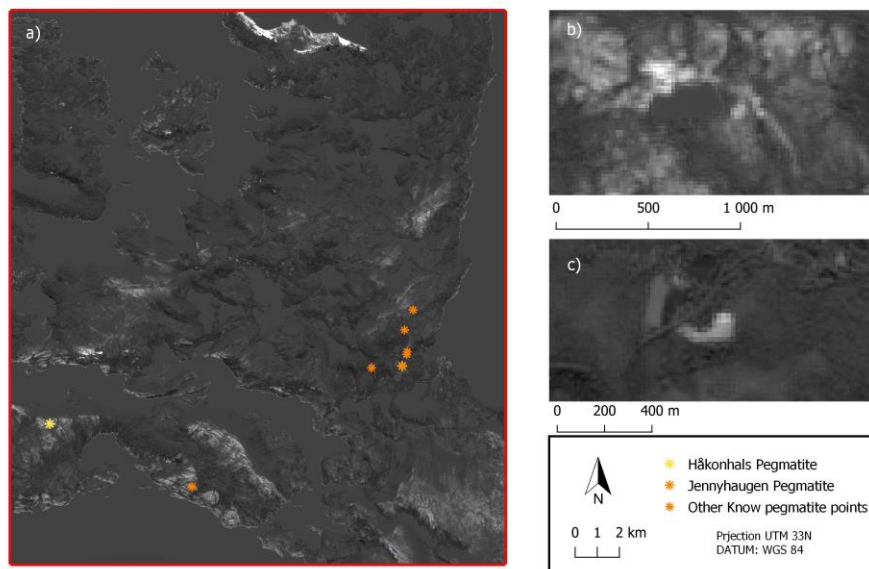


Figure 27. PC 2 bands of the bands 8,12 for Sentinel-2.

3.1.2. Band Ratios

Some of the band ratios tested, such as 2/12, 2/3, 3/4, 6/3 and others did not highlight the pixels of the Håkonhals and Jennyhaugen mines very well. In these ratios there was a lot of signal confusion with water or granite and, as in the case of the 5/12 ratio, the image resolution impairs the analysis. Because of this, they will not be presented individually in this study. The 8/12 and 7/12 ratios, although not highlighting the mines, detect the granitic zones in dark pixels very well. The ratio 3/8 (figure 28), highlighted the Jenny-Haugen mine better. This may be due to the differences in the minerals extracted from the two mines. The ratio 4/8 (figure 29), was the ratio that best highlighted the two mines, as the previous one highlighted the pixels of interest in light grey. The ratio 4/6 (figure 30), showed strong noise in the area of the Håkonhals mine but was nevertheless able to identify the Jennyhaugen mine well highlight the mine pixels in light grey, but still also highlight the water with pixels of the same color which is not ideal.

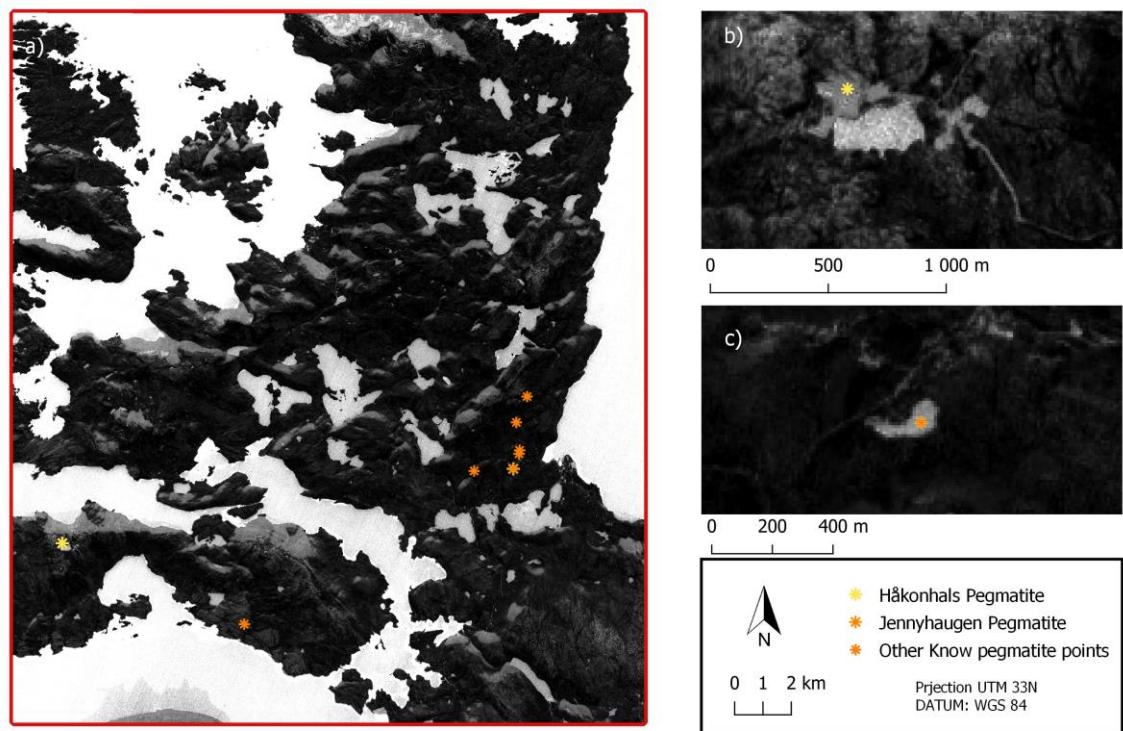


Figure 28. Band Ratios 3/8 for Sentinel-2. The pixels of the mines are highlighted in light grey.

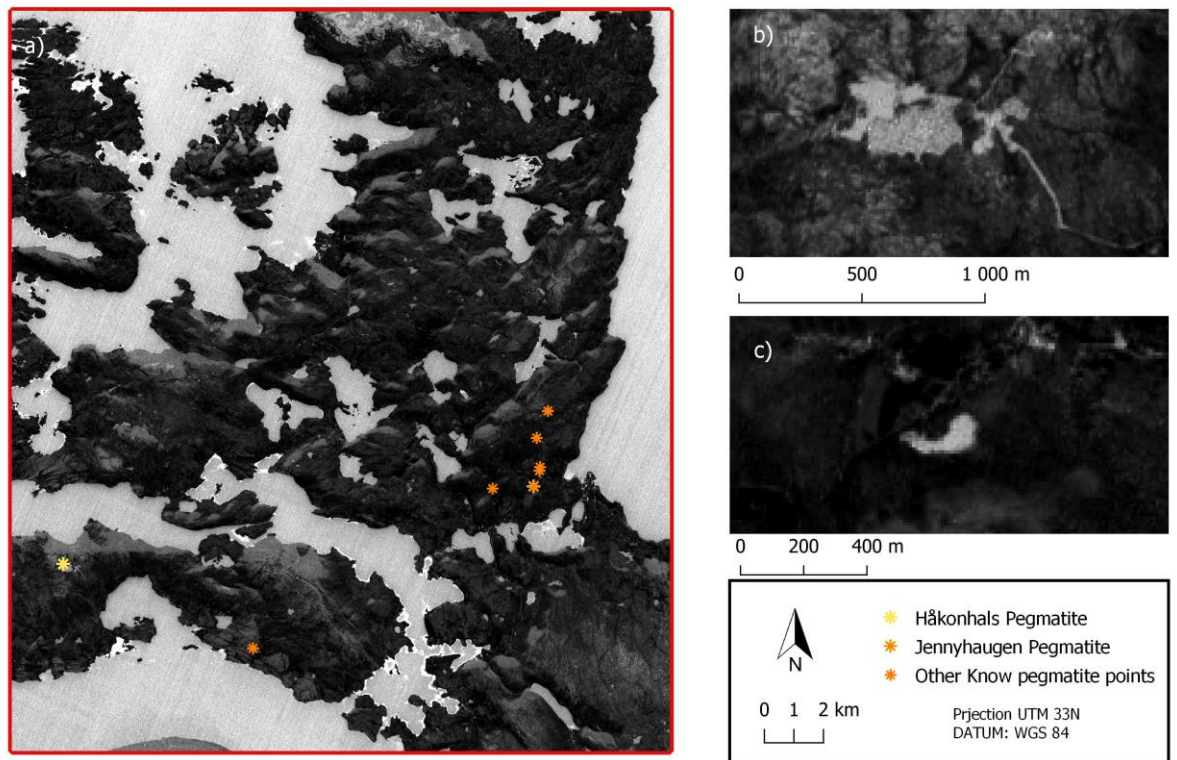


Figure 29. Band Ratios 4/8 for Sentinel-2. The pixels of the mines are highlighted in light grey.

Source: By Author adapted from USGS.

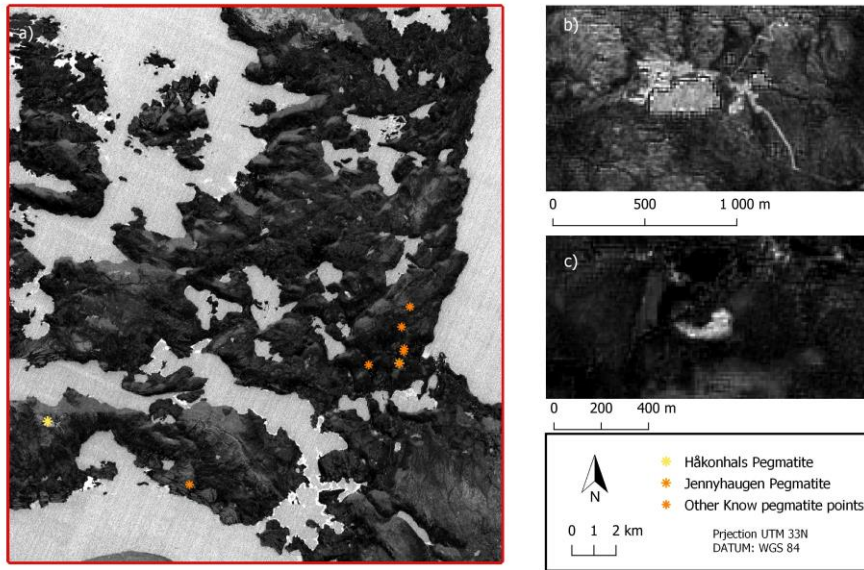


Figure 30. Band Ratios 4/6 for Sentinel-2. The pixels of the mines are highlighted in light grey.

The divisions made by bands belonging to the same spectrum such as 2/4, 2/3, 4/3, 7/6, 8/6, 11/12, did not obtain acceptable results. On the other hand, the best results among the BR, which were able to highlight, even minimally, the known pegmatite areas, were the visible (reflecting) bands divided by an absorbing NIR or SWIR band (example BR 2/8, 3/8 and 2/12).

Following the example of what was done by Simon *et al.* (2016), ratios were inserted into RGB fields to form new RGB combinations of ratios. The combination 4/8, 2/8 and 8/12 (figure 31) highlights the areas of mines in orange, water in yellow, granite in dark orange color and vegetated areas are highlighted in blue color. The combination of ratios 4/8, 3/8 and 8/12 (figure 32), highlights the areas of mines in orange, water in yellow, granite in dark orange color and vegetated areas in blue color. The combinations 4/8, 2/8 and 8/12 and 4/8, 3/8 e 8/12 present signal confusion with a granite outcrop that has been highlighted in the same color as the mine. This particularity should be studied through fieldwork. The combinations 4/8, 8/4 and 8/12 (figure 33), highlight the mining areas in red, the areas with granite in light red and the vegetation in light green. This combination presented both signal confusion with the water and granitic outcrop.

Individually the BRs tested do not have very significant results regarding the detection of NYF pegmatites, but when they are combined in RGB compositions, they prove to be powerful resources for the identification of these pegmatites.

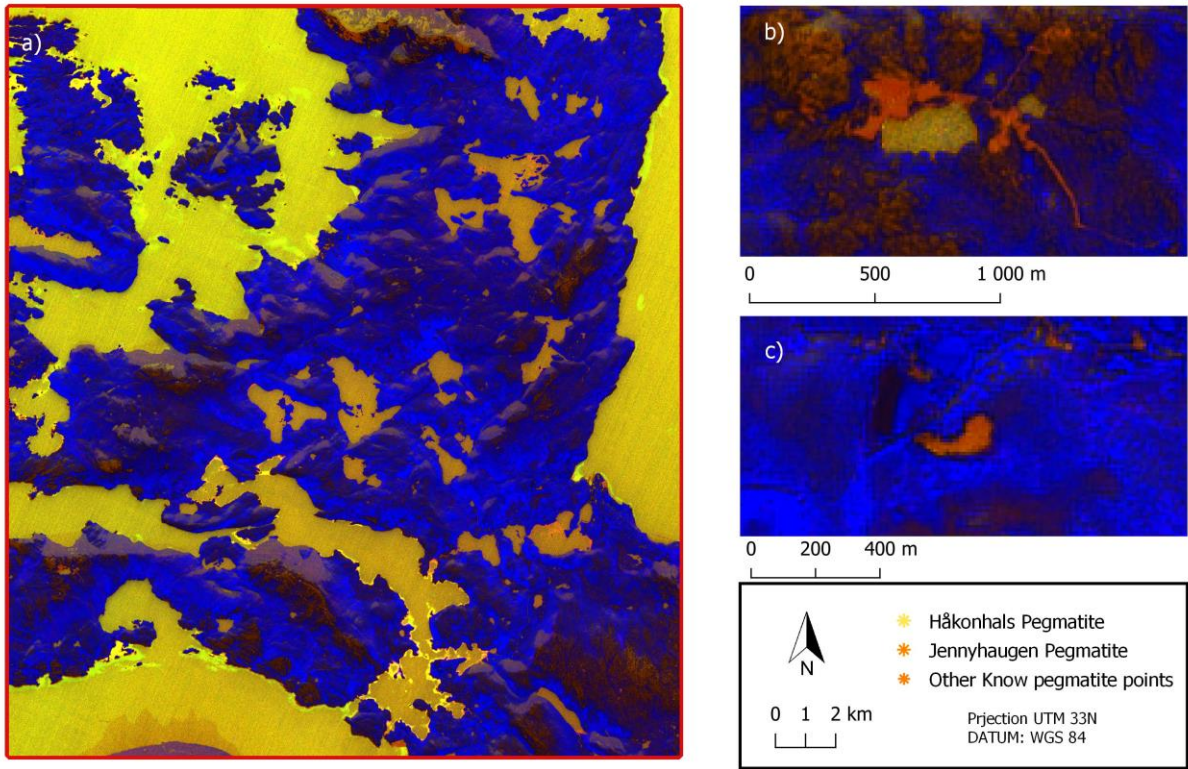


Figure 31. Combination of ratios 4/8, 2/8 and 8/12 in RGB composite

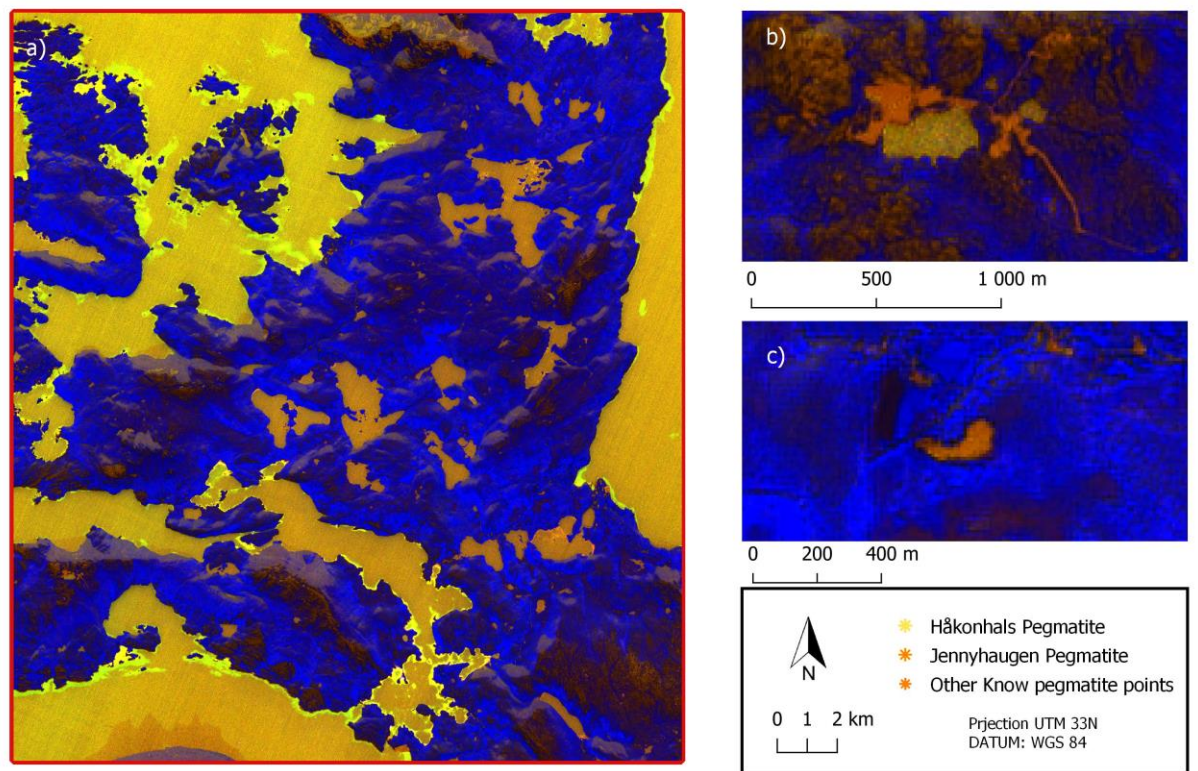


Figure 32. Combination of ratios 4/8, 3/8 and 8/12 in RGB composite.

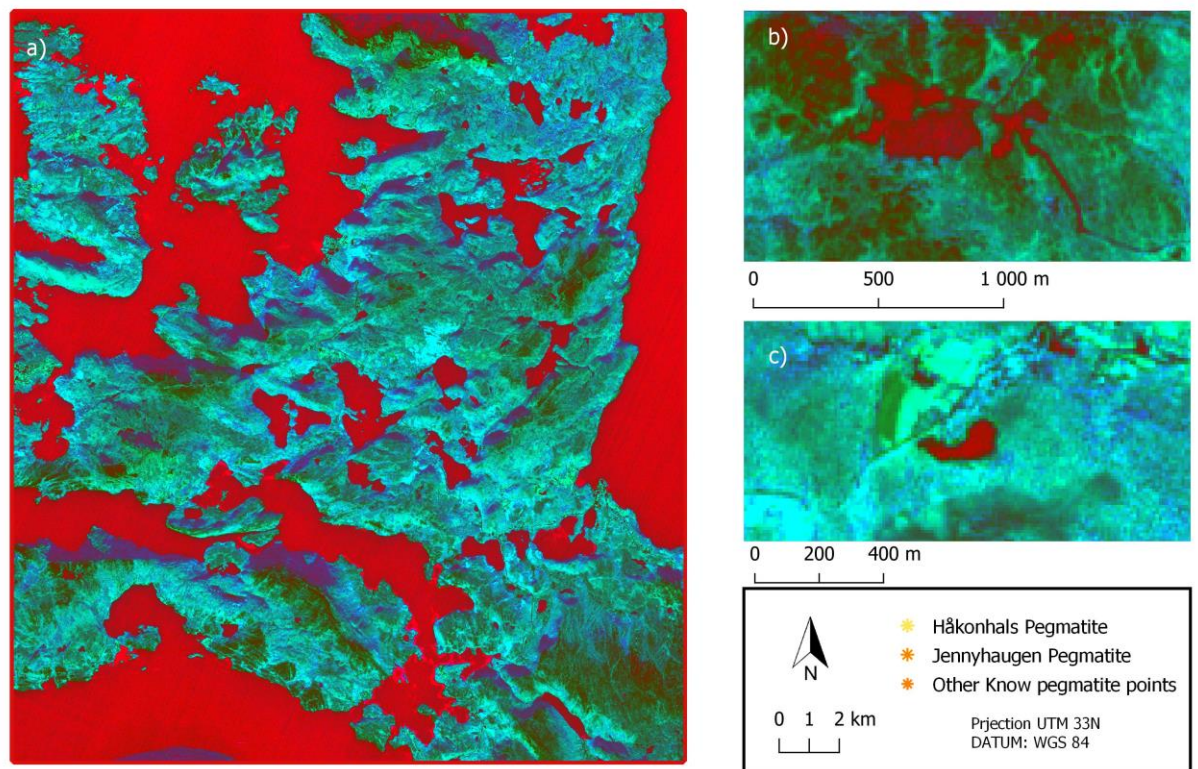


Figure 33. Combination of ratios 4/8, 8/4 and 8/12 in RGB composite.

3.2. Third set: Machine Learning algorithms

The algorithms were tested with several variables, of these the one that worked best in both algorithms was the "balanced" variable with the "region of interest (ROI)" method to split the data.

3.2.1. Results and comparasion

As shown in Table 8, both algorithms have very high socre values (test, training and OA), but this does not necessarily mean something positive because, these high values indicate overfitting of the models. Even though they showed very close accuracy value (0.963 for SVM and 0.961 for RF), SVM, in general, shows more acceptable values for train test results.

Table 8. 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

Regarding classification, both obtained similar results concerning the identification of pegmatite bodies, managing to highlight five known pegmatite areas in the Tysfjord district (figure 34). The figure 35, show in more detail the identification of the five areas and how both algorithms performed well in identifying the target (pegmatite bodies). When we analyze the figure 36, comparatively, is perceptible that the RF algorithm classified a greater amount of pixels as granite and the SVM algorithm, on the other hand, highlighted more pixels as water bodies and vegetation.

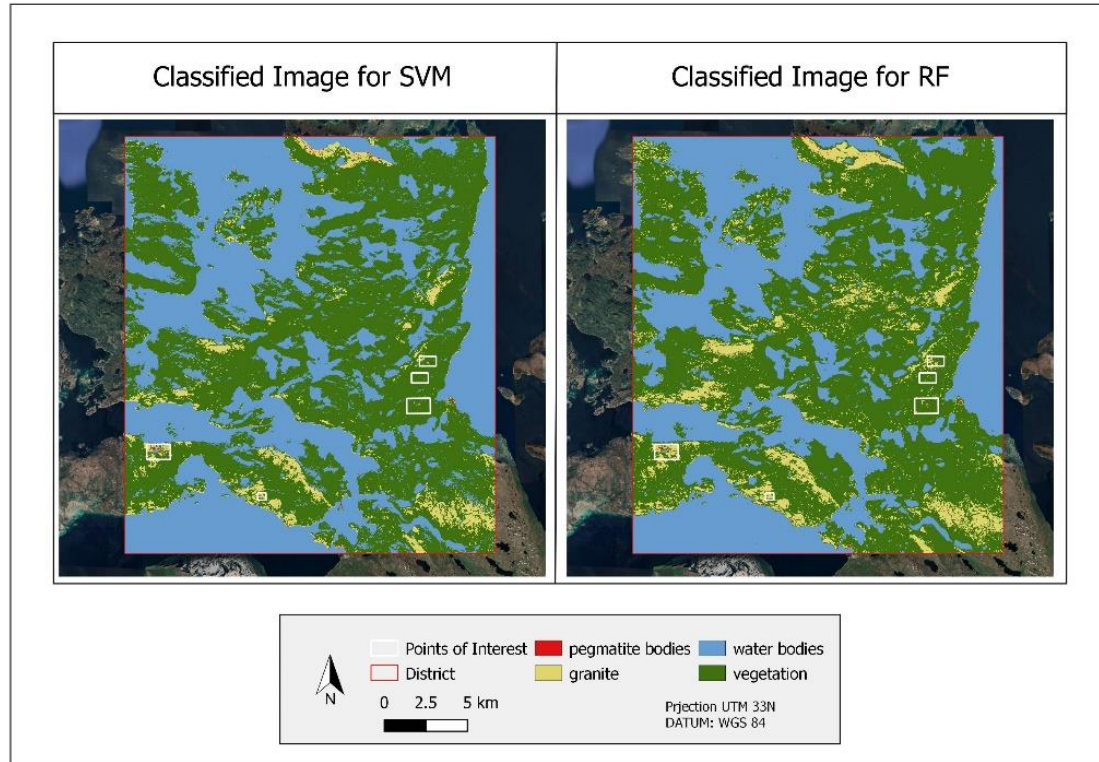


Figure 34. Comparison of results for the SVM (right) and RF (left) algorithm classifications.

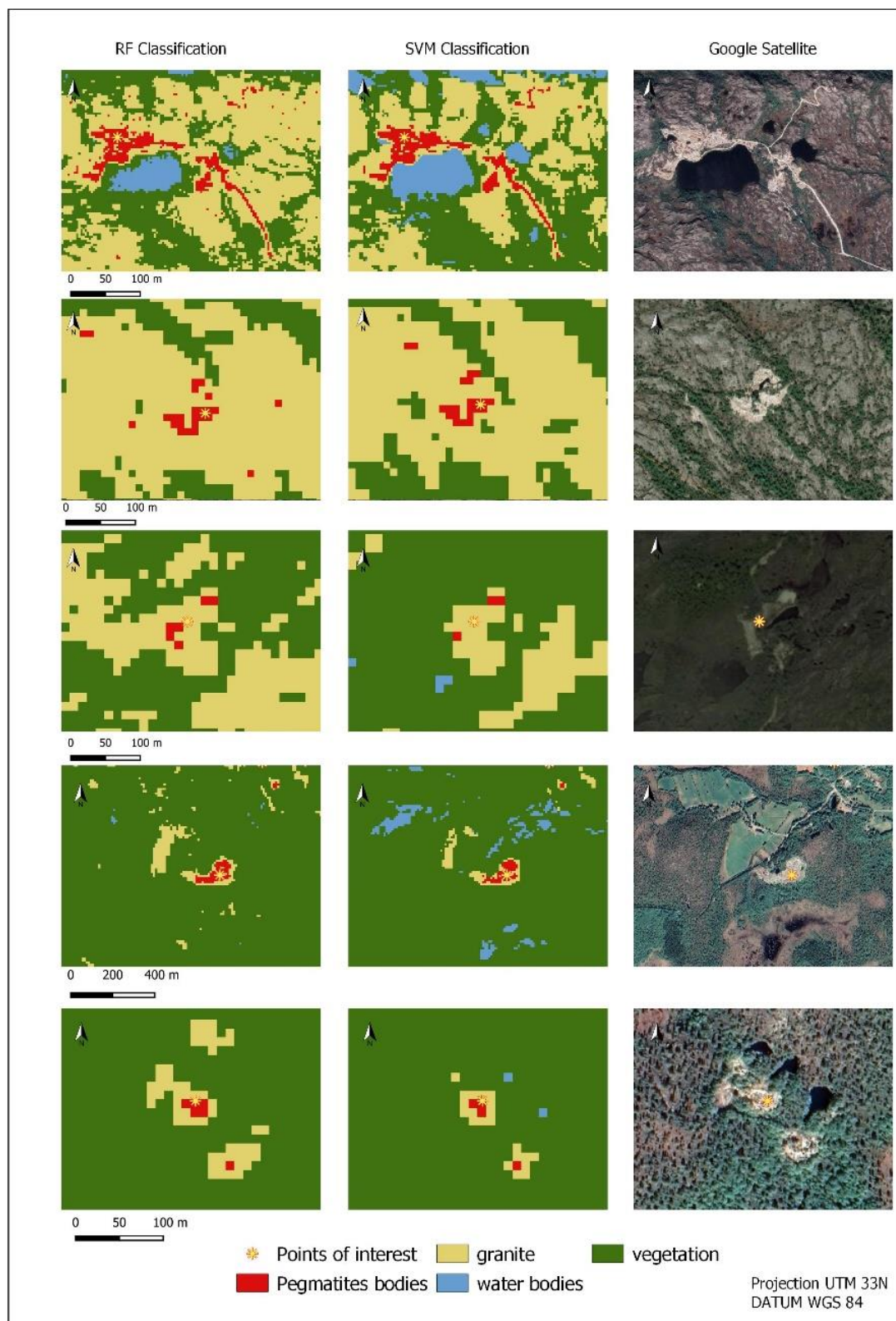


Figure 35. Comparison of the classification of the two algorithms for the known pegmatite areas.

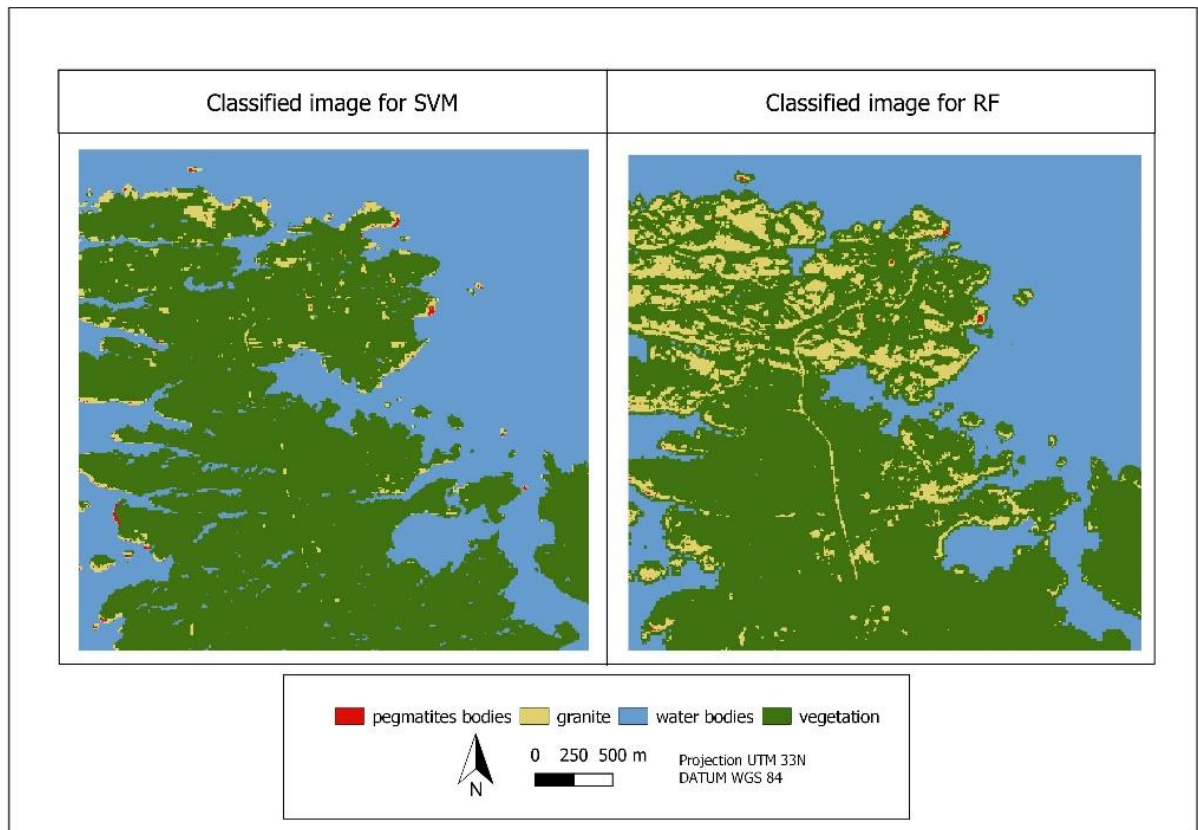


Figure 36. Exemplification of the difference in granite classification between the two algorithms.

When comparing the results of the two algorithms, it is noted that RF had a better result in the classification of the granitic outcrop, identifying more accurately the pixels of this class. On the other hand, the RF algorithm shows a lower sensitivity to the spectral differences between the classes, assuming and misclassifying the pixels as we can see in figure 37, where it can be observed that the RF algorithm confused the signals of vegetation with water. Although it did not classify the granitic areas very well, classifying most of the granitic outcrops as vegetation, the SVM algorithm resulted in a more assertive classification and a better separation between classes which demonstrates that it was more sensitive to spectral differences.

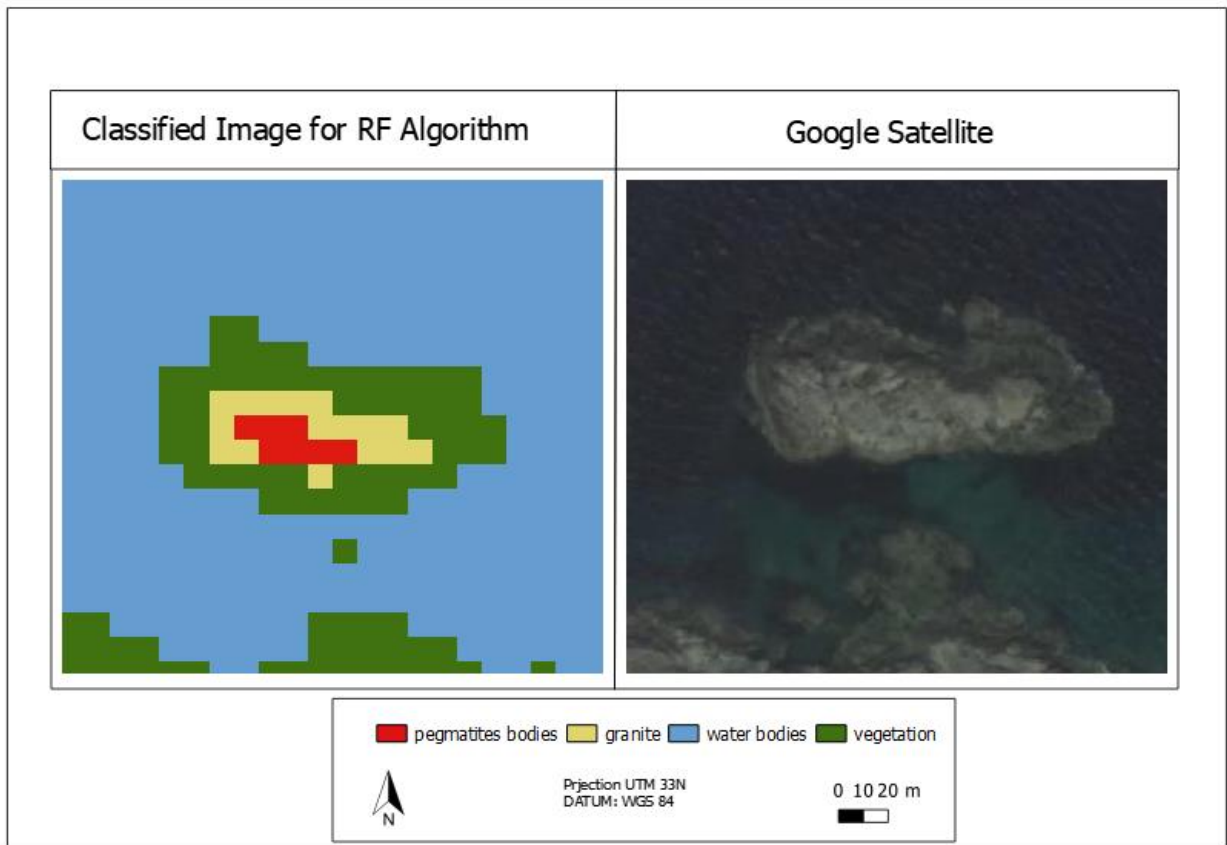


Figure 37. Comparison between RF classification and Google satellite image where RF assumes as vegetation pixels that correspond to water.

After analyzing the results and based on patterns that the algorithm identified in known pegmatite areas, four new areas of possible interest were highlighted (figure 38). These four areas are:

point a) located at coordinates 536401.86 m E/ 7556336.90 m N, where RF highlighting a larger number of area as pegmatites (red color) than SVM;

point b) located at coordinates 528301.13 m E/ 7559019. 79 m N, corresponds to an area on the side of a dirt road, which unlike the previous one, had a greater number of highlighted pixels in the SVM classification;

point c) located at coordinates 543539.60 m E / 7548600. 89 m N, near three points of pegmatites already known, which may indicate a possible deposit of tailings from these mines, with a higher number of pixels classified as pegmatites with the RF algorithm; and;

point d) at coordinates 527454.93 m E / 7556955.90 m N, having obtained more prominence (higher number of pixels) with the SVM algorithm.

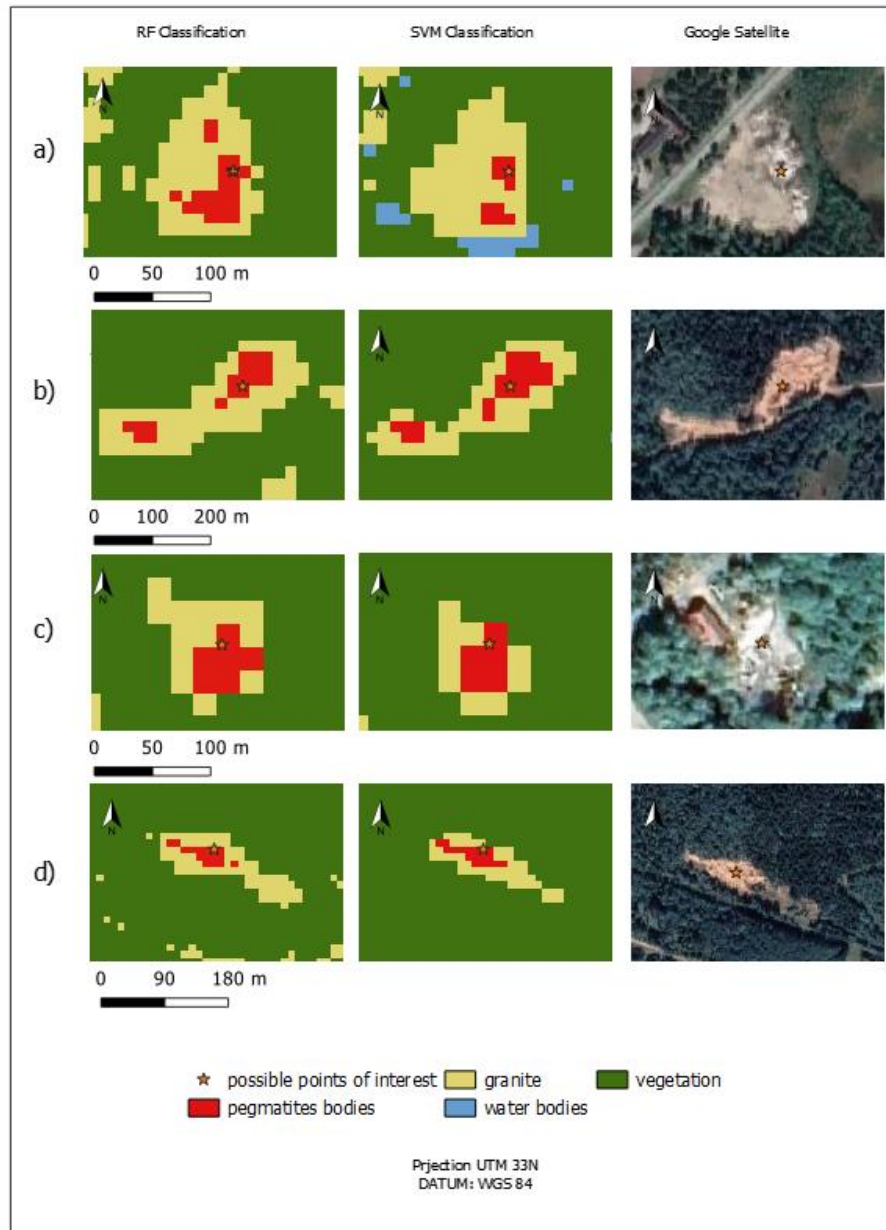


Figure 38. Comparison of the prominence of possible areas with the presence of pegmatites

As we can observe in figure 39, in relation to false positives, we can verify that a good part is generated by the confusion of signals with buildings and coastal areas, such as intertidal/supradital areas and, to a lesser extent, with steep slopes.

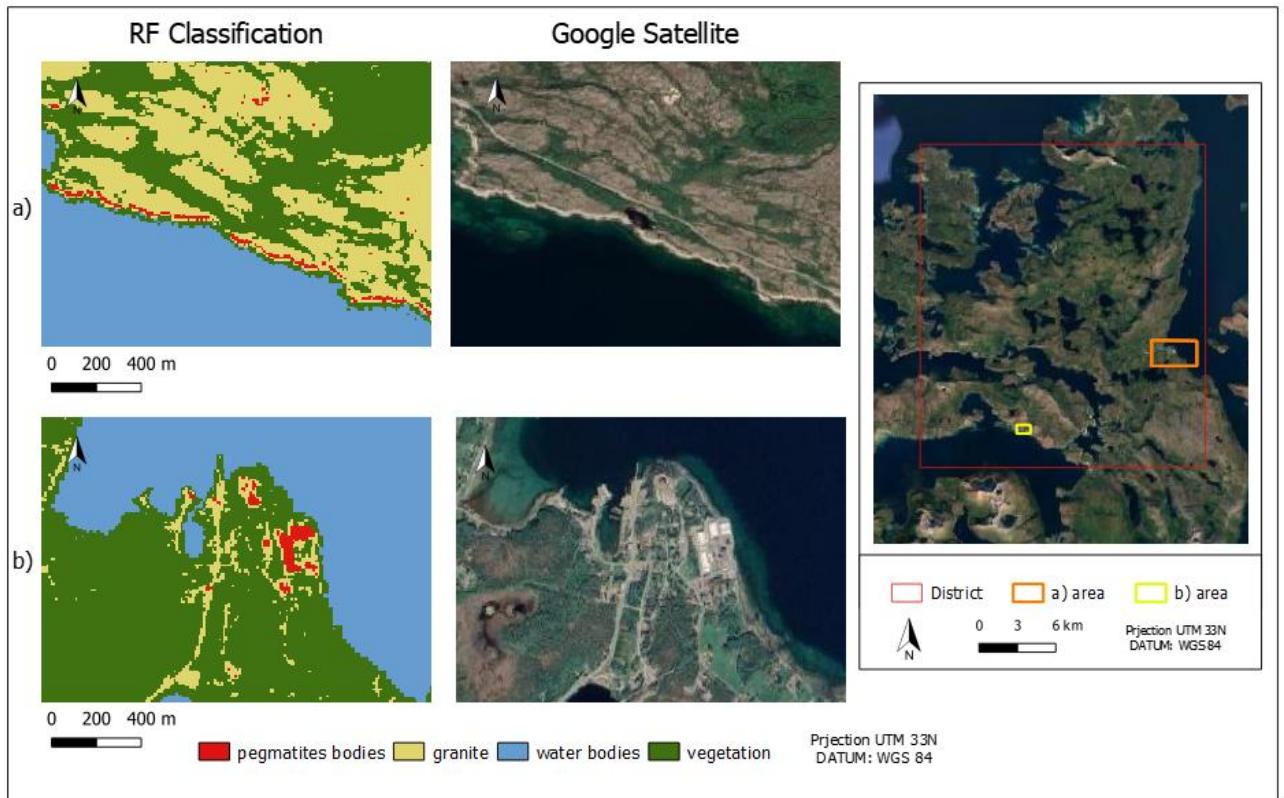


Figure 39. Examples of false positives obtained with the RF algorithm. Source: By Author adapted from USGS and Google Earth. a) an example of the signal confusion in a coastal area; b) an example of the signal confusion between buildings and pegmatites.

When we compare the results of all the methods (figure 40), we can see that although the best BR and PC were able to identify the known pegmatite areas, they still do not compare to the result obtained by the ML algorithms. These algorithms, besides being able to identify the known pegmatite areas, also presented a small amount of false positives.

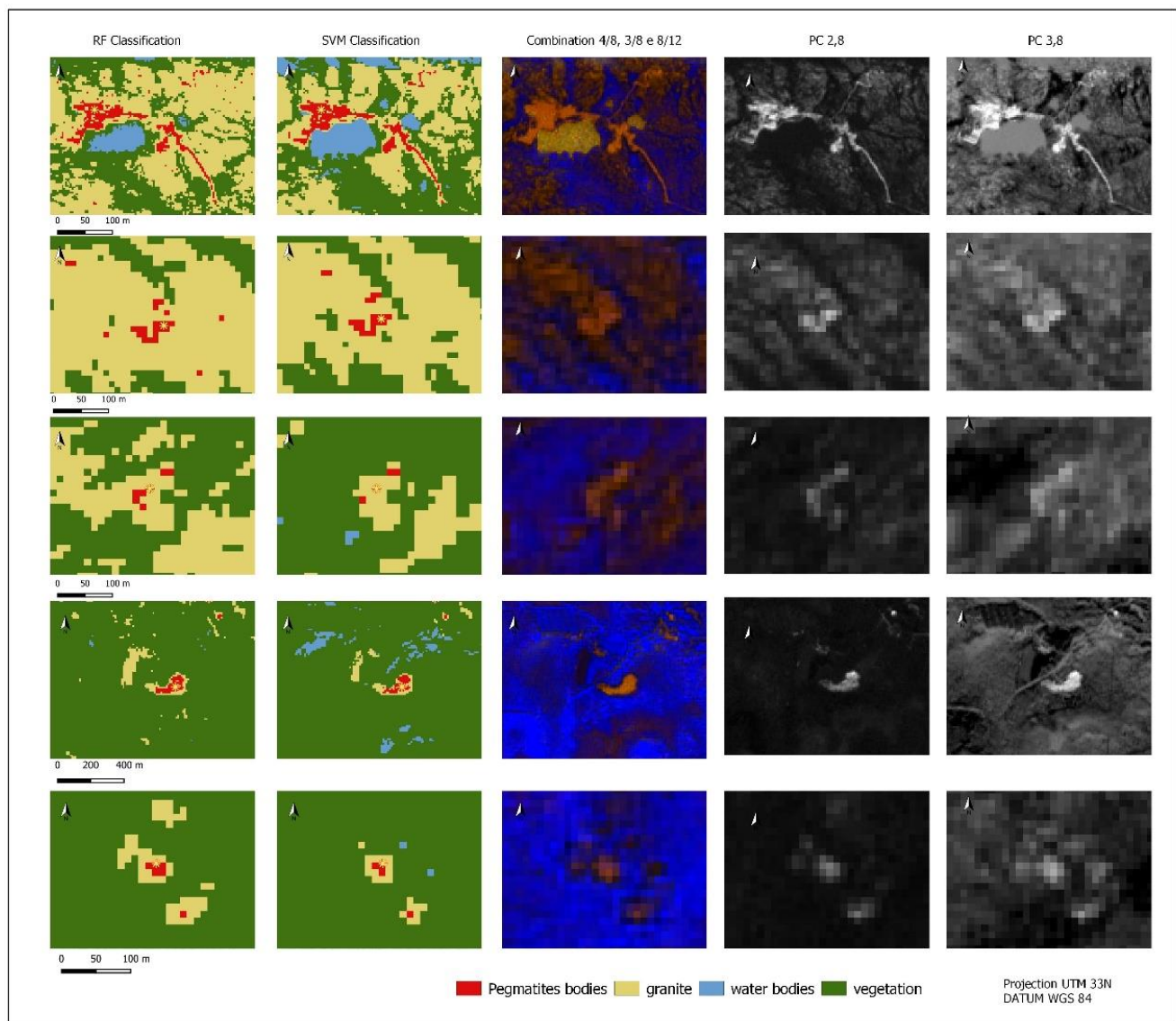


Figure 40. Comparison of the best results of each set.

4. Conclusions

Despite in the first set: “Results for LCT traditional methods”, only one of the methods tested was able to identify the know pegmatites bodies (corresponding to the PC of bands 3.8 for Sentinel-2). The fact that this method highlights known pegmatite mines so well attests that there are methods for identifying LCT pegmatites that also work in identifying NYF-type pegmatites.

In the second set, we can attest that, individually, the BR did not show satisfactory results, but when combined to RGB combinations were able to obtain encouraging results regarding the detection of NYF type pegmatites. The ratios 4/8, 2/8 and 8/12 and 4/8, 3/8 and 8/12 present the same result. The other ratios present noise or too much interference from other elements of the study area such as water and granites outcrops.

Regarding the ML algorithms (third set), this method showed to be the most efficient and powerful method to identify pegmatite bodies. The pegmatite area was identified for both algorithms (SVM and RF), and despite the signal confusion with coastal areas and constructed areas like step lops, these two algorithms showed better results than traditional methods tested. It is worth mentioning that the false positives were decreasing as the algorithm was being improved. When we analyze the results of the RF algorithm, we can see that it was affected by the high overfitting, which may explain the misclassification between vegetation and water bodies. Regarding the ML algorithms tested, we can say that the SVM algorithm on the variable "balanced" with the "region of interest (ROI)", obtained a better performance, in general, in the image classification and both algorithms obtained similar results regarding the identification of the sample areas (known pegmatite areas).

Although the ML algorithms have obtained better results than the other methods tested, their applicability (despite the equal accessibility of free multispectral images) is more complex and, even consulting the literature on the subject, requires more specific and in-depth knowledge of programming languages. The other methods, on the other hand, have less complex applicability, which makes them more accessible and practical, this being their main advantage over the ML employed in this thesis.

The results of this thesis are a strong contribution to the field of remote sensing applied to pegmatite identification. The methods applied in the second set can be evaluated, improved and refined with spectral analysis of further samples from the study area. As was done in the second

set, traditional methods can be improved and adapted to be applied in areas with different geological characteristics and to identify other targets such as other minerals that may be of interest. The spectral analysis of the Tysfjord samples was essential for the adaptation of the traditional methods, being the necessary component to make these methods adaptable to new areas of study.

From the work developed in this thesis a conference proceeding entitled "Identification of pegmatite bodies, at a province scale, using machine learning algorithms: preliminary results" was presented in SPIE Europe Remote Sensing and published (Teodoro et al., 2021). This proceeding is attached (Appendix).

4.1. Future work

The covid 19 pandemic, caused a huge delay in this research by preventing the collection of samples in the field. The samples sent by project partners were sufficient for the analyses performed in this work, but more samples collected not only from existing points but also from new points of interest should be collected in the future and used to validate will make this research more robust.

The ML algorithms although powerful presented high overfitting in their model, even so, they still presented great results which suggests that by decreasing the overfitting we can make these results even better. Working on the implementation and improvement of the ML model, adding and testing new inputs in its composition is a necessary task.

References

- AMER, R.; MEZAYEN, A. E. L.; HASANEIN, M. - ASTER spectral analysis for alteration minerals associated with gold mineralization. *Ore Geology Reviews*. ISSN 0169-1368. Vol. 75 (2016), p. 239–251. doi: <https://doi.org/10.1016/j.oregeorev.2015.12.008>
- BAGHDADI, N.; ZRIBI, M. – Introduction. In *Optical Remote Sensing of Land Surface: Techniques and Methods* [Em linha]. Oxford: Elsevier, 2017. [4 May 2021]. Available at: <https://www.sciencedirect.com/book/9781785481024/optical-remote-sensing-of-land-surface> ISBN 978-1-78548-102-4. doi: 10.1016/B978-1-78548-102-4.50012-0
- BAKKER, W. H.; FERINGA, W.; GIESKE, A. S. M.; [*et al.*] - Principles of Remote Sensing - An introductory textbook. [Em Linha] Netherlands: The International Institute for Geo-Information Science and Earth Observation (ITC), 2009. [02 Jun 2021]. Available at: https://webapps.itc.utwente.nl/librarywww/papers_2009/general/principlesremotesensing.pdf ISBN 978–90–6164–270–1
- BALDWIN, J. R.; KNORRING, O. V. - Compositional range of Mn-garnet in zoned granitic pegmatites. *Canadian Mineralogist*. ISSN 00084476. Vol. 21:4 (1983), p. 683–688.
- BARROSO, L. A.; RUDRICK, H.; SENSFUSS, F.; LINARES, P. - The green effect. *IEEE Power and Energy Magazine*. ISSN: 1540-7977. Vol. 8:5 (2010), p. 35. doi: 10.1109/MPE.2010.937595
- BISWAJEET, P.; SARO, L. - Utilization of Optical Remote Sensing Data and GIS Tools for Regional Landslide Hazard Analysis Using an Artificial Neural Network Model. *Earth Science Frontiers*. ISSN 18725791. Vol. 14:6 (2007), p. 143–151. doi: 10.1016/s1872-5791(08)60008-1
- BLENGINI, G.A.; LATUNUSSA, C.E.; EYNARD, U.; WITTMER, D.; GEORGITZIKIS, K.; PAVEL, C.; CARRARA, S.; MANCINI, L.; UNGURU, M.; BLAGOEVA, D.; MATHIEUX, F.; PENNINGTON, D. - Study on the EU's list of Critical Raw Materials (2020): Final Report. [Em linha]. Luxembourg: Publications Office of the European Union, 2020.

[16 april 2021]. Available at: < <https://op.europa.eu/en/publication-detail/-/publication/8dabb4c1-f894-11ea-991b-01aa75ed71a1/language-en>>

BOBBA, S.; CARRARA, S.; HUISMAN, J.; MATHIEUX, F.; Pavel, C. - Critical Raw Materials for Strategic Technologies and Sectors in the EU: a Foresight Study [Em linha]. Luxembourg: Publications Office of the European Union, 2020. [21 May 2021]. Available at:<URL:<https://ec.europa.eu/docsroom/documents/42881>>. ISBN 9789276153375

BYRNE, G. F.; CRAPPER, P. F.; MAYO, K. K. - Monitoring land-cover change by principal component analysis of multitemporal landsat data. Remote Sensing of Environment. ISSN 00344257. Vol. 10:3 (1980), p. 175–184. doi: 10.1016/0034-4257(80)90021-8

CARDOSO-FERNANDES, J.; TEODORO, A. C.; LIMA, A. - Potential of Sentinel-2 data in the detection of lithium (Li)-bearing pegmatites: a study case. Proc. SPIE - Earth Resources and Environmental Remote Sensing/GIS Applications IX. ISSN 1996756X. Vol. 10790 (2018^a), p. 201-215. doi: 10.1117/12.2326285

CARDOSO-FERNANDES, J. - Remote sensing applied to the Bajoca lithiniferous area. Porto: Faculdade de Ciências da Universidade do Porto (FCUP), 2018^b. p. 199. Master's thesis.

CARDOSO-FERNANDES, J.; TEODORO, A. C.; LIMA, A.; RODA-ROBLES, E.- Evaluating the performance of support vector machines (SVMs) and random forest (RF) in Li-pegmatite mapping: preliminary results. Proc. SPIE - Earth Resources and Environmental Remote Sensing/GIS Applications X. ISSN 1996756X. Vol. 11156 (2019^b), p. 146-157. doi: 10.1117/12.2532577

CARDOSO-FERNANDES, J.; LIMA, A.; RODA-ROBLES, E.; TEODORO, A. C. - Constraints and potentials of remote sensing data/techniques applied to lithium (Li)-pegmatites. Canadian Mineralogist. ISSN 00084476. Vol. 57:5 (2019^b), p.723–725. doi: 10.3749/canmin.AB00004

CARDOSO-FERNANDES, J.; TEODORO, A. C.; LIMA, A.; PERROTTA, M.; RODA-ROBLES, E. - Detecting Lithium (Li) mineralizations from space: Current research

- and future perspectives. *Applied Sciences* (Switzerland). ISSN 20763417. Vol. 10:5 (2020^a), p. 1785. doi: 10.3390/app10051785
- CARDOSO-FERNANDES, J.; TEODORO, A. C.; LIMA, A. - Remote sensing data in lithium (Li) exploration: A new approach for the detection of Li-bearing pegmatites. *International Journal of Applied Earth Observation and Geoinformation*. ISSN 1872826X. Vol. 76 (2019^a), p.10–25. doi: 10.1016/j.jag.2018.11.001
- CARDOSO-FERNANDES, J.; TEODORO, A. C. M.; LIMA, A.; RODA-ROBLES, E.- Semi-Automatization of Support Vector Machines to Map Lithium (Li) Bearing Pegmatites. *Remote Sensing*. ISSN 2072-4292. Vol 12:14 (2020^b), p. 22. doi: 10.3390/rs12142319
- CARDOSO-FERNANDES, J.; SILVA, J.; DIAS, F.; Lima, A.; Teodoro, A. C.; Barrès, O.; Cauzid, J.; Perrotta, M.; Roda-robles, E.; Ribeiro, M. A. - Tools for remote exploration: A lithium (li) dedicated spectral library of the fregeneda–almendra aplite–pegmatite field. *Remote Sensing*. ISSN 23065729. Vol. 6:3 (2021), p. 10. doi: 10.3390/data6030033
- CERASOLI, S.; CAMPAGNOLO, M.; FARIA, J.; Nogueira, C.; Da Conceição - Caldeira, M. - On estimating the gross primary productivity of Mediterranean grasslands under different fertilization regimes using vegetation indices and hyperspectral reflectance. *Biogeosciences*. ISSN 17264189. Vol. 15:17 (2018), p. 5455–5471. doi: 10.5194/bg-15-5455-2018
- CERNY, P. - Rare-element granitic pegmatites. Part I: anatomy and internal evolution of pegmatite deposits. *Geoscience Canada* . ISSN 03150941. Vol. 18:2 (1991), p. 49–67.
- ČERNÝ, P.; ERCIT, T. S. - The classification of granitic pegmatites revisited. *Canadian Mineralogist*. ISSN 00084476. Vol. 43:6 (2005), p. 2005–2026. doi: 10.2113/gscanmin.43.6.2005.
- CHAVEZ JR, P. S. – Image Based Atmospheric Corrections: Revisited and Improved. *Remote Sensing*. ISSN 0099-1112. Vol. 62 (1996), p. 1025–1036.

- CHAKHMOURADIAN, A.; SMITH, M.; KYNICKY, J. - From “strategic” tungsten to “green” neodymium: A century of critical metals at a glance. *Ore Geology Reviews*. ISSN 0169-1368. Vol. 64 (2015), p. 455–458. doi: 10.1016/j.oregeorev.2014.06.008
- CHEN, D. M.; STOW, D. - The effect of training strategies on supervised classification at different spatial resolutions. *Photogrammetric Engineering and Remote Sensing*. ISSN 00991112. Vol. 68:11 (2002), p. 1155–1161.
- EASTMAN JR. - IDRISI for Windows: Introdução e exercícios tutoriais. Centro de Recursos Idrisi Brasil. 2nd ed. Porto Alegre: Universidade Federal do Rio Grande do Sul (UFRGS), 1998.
- EUROPEAN COMMISSION - Report on critical raw materials for the EU: Report of the Ad hoc Working Group on defining critical raw materials. [Em linha]. Luxembourg: Publications Office of the European Union, 2014. [23 May 2021]. Available at: <URL:http://ec.europa.eu/enterprise/policies/raw-materials/files/docs/crm-report-on-critical-raw-materials_en.pdf>.
- FITZ, P. R.; VIEIRA, J. C.; SOARES, M. C. - Use of sampling polygons in supervised classifications of satellite images. *ENTRE-LUGAR*. ISSN 0103-8427. Vol. 10:19 (2019), p. 319–341. doi: 10.30612/el.v10i19.9595
- FOSLIE, S. - Tysfjords geologi. Beskrivelse til det geologiske gradteigskart Tysfjord. *Angewandte Chemie International* [em linha]. Oslo: NORGES GEOLOGISKE UNDERSØKELSE, 1941. [15 April 2021]. available at: <<https://openarchive.ngu.no/ngu-xmlui/handle/11250/2674532>>
- GAMALETSOS, P. N.; GODELISTAS, A.; FILIPPIDIS, A.; PONTIKES, Y. - The Rare Earth Elements Potential of Greek Bauxite Active Mines in the Light of a Sustainable REE Demand. *Journal of Sustainable Metallurgy*. ISSN 21993831. Vol. 5:1 (2019), p. 20–47. doi: 10.1007/s40831-018-0192-2.
- GEMUSSE, U.; LIMA, A.; TEODORO, A. C. M. - Comparing different techniques of satellite imagery classification to mineral mapping pegmatite of Muiane and Naipa: Mozambique. *Proc. SPIE - Earth Resources and Environmental Remote Sensing/GIS*

- Applications X. ISSN 1996756X. Vol. 111561E (2019), p. 49. doi: 10.1117/12.2532570.
- GERON, A. - Hands-On Machine Learning with Scikit: Learn and TensorFlow. Sebastopol: O'Reilly Media, 2019. ISBN 9781492032649
- Geomatics 2018 - [Em linha]. [12 set 2020]. available at :<URL:www.pcigeomatics.com>.
- GRÜBLER, A.; NAKIĆENOVIĆ, N. - Decarbonizing the global energy system. Technological Forecasting and Social Change. ISSN 00401625. Vol. 53:1 (1996), p. 97–110. doi: 10.1016/0040-1625(96)00049-2
- GISLASON, P. O.; BENEDIKTSSON, J. A.; SVEINSSON, J. R. - Random forests for land cover classification. Pattern Recognition Letters. ISSN 01678655. Vol. 27:4 (2006), p. 294–300. doi: 10.1016/j.patrec.2005.08.011
- GOETZ, A.F. H.; ROWAN, L. C.; KINGSTON, M. J. - Mineral Identification from Orbit: Initial Results from the Shuttle Multispectral Infrared. Science. ISSN 0036-8075. Vol. 218 (2016), p. 1020–1024. DOI: 10.1126/science.218.4576.1020
- GONZALEZ, R. C.; WINTZ, P. - Digital image processing. In Applied Mathematics and Computation. 6th ed. Illustrated. University of California: Addison-Wesley Publishing Co., Inc, 1977. ISBN 13: 9780201030457
- GONZÁLEZ, B.; FREIRE, I.; MORENTE, C. F.; PÉREZ ASENSIO, E.- Los sistemas de información geográfica y la investigación en ciencias humanas y sociales [Em linha]. Madrid : Confederación Española de Centros de Estudios Locales, 2012. [2 Jun 2021] Available at: <URL:http://hdl.handle.net/10261/64940>. ISBN 9788461598250
- HARRIS, J. R.; WILKINSON, L.; GRUNSKY, E. C. - Effective use and interpretation of lithogeochemical data in regional mineral exploration programs: Application of Geographic Information Systems (GIS) technology. Ore Geology Reviews. ISSN 01691368. Vol. 16:3–4 (2000), p. 107–143. doi: 10.1016/S0169-1368(99)00027-X
- HAZEN, R. M.; PAPINEAU, D.; BLEEKER, W.; DOWNS, R. T.; FERRY, J. M.; MCCOY, T. J.; SVERJENSKY, D. A.; YANG, H.- Mineral evolution. American Mineralogist. ISSN 0003004X. Vol. 93:11-12 (2008), p. 1693–1720. doi: 10.2138/am.2008.2955

- HAXEL, G. B.; HEDRICK, J. B.; ORRIS, G. J.; STAUFFER, P. H.; HENDLEY II, J. W. - Rare Earth Elements: Critical Resources for High Technology. [Em Linha], 2002. [17 May 2021] available at: <URL:<http://pubs.er.usgs.gov/publication/fs08702>>
- HEINRICH, E. W. - Pegmatites of Eight Mile Park, Fremont County, Colorado. American Mineralogist. ISSN 0003-004X. Vol. 33:7–8 (1948), p. 420–448.
- HETHERINGTON, C. J.; MAILLOUX, G. A.; 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. ISSN 18726143. Vol. 398–399 (2021), p. 18. doi: 10.1016/j.lithos.2021.106257
- HUSDAL, T. - The minerals of the pegmatites within the Tysfjord granite, northern Norway. Norsk Bergverksmuseum skrift. ISSN 0800-1855. Vol. 38:1906 (2008), p. 5–28.
- JHA, A.; TUKKARAJA, P. - Monitoring and assessment of underground climatic conditions using sensors and GIS tools. International Journal of Mining Science and Technology. ISSN 20952686. Vol. 30:4 (2020), p. 495–499. doi: 10.1016/j.ijmst.2020.05.010
- KESLER, S. E.; GRUBER, P. W.; MEDINAC, P. A.; Keoleian, G. A.; Eversond, M. P.; Wallington, T. J. - Global lithium resources: Relative importance of pegmatite, brine and other deposits. Ore Geology Reviews. ISSN 01691368. Vol. 48 (2012), p. 55–69. doi: 10.1016/j.oregeorev.2012.05.006
- LONDON, D. - Pegmatites. In Encyclopedia of Geology. 2nd ed. Oklahoma: Elsevier, 2021. ISBN 9780124095489. p. 184–195 . doi: 10.1016/B978-0-12-409548-9.12489-3.
- LOUGHLIN, W. P. - Principal Component Analysis for Alteration Mapping. Photogrammetric Engineering & Remote Sensing. ISSN: 0099-1112. Vol. 57:9 (1991), p. 1663–1669.
- MACHADO, A. R. J. - A emergência dos sistemas de informação geográfica na análise e organização do espaço. Lisboa: Fundação Calouste Gulbenkian, 2000. ISBN 9789723108187.

- MARTIN, D. - Geographic information systems: socioeconomic applications. London: Routledge, 1996. ISBN: 0415125723.
- MARTIN, M.; FOX, R. V.; Miziolek, A. W.; DeLucia, F. C.- Spectral analysis of rare earth elements using laser-induced breakdown spectroscopy, Proc. SPIE – Next Generation Spectroscopic Technologies VIII. ISSN 0277-786X. Vol. 9482 (2015). doi:10.1117/12.2178192
- MASSARI, S.; RUBERTI, M. - Rare earth elements as critical raw materials: Focus on international markets and future strategies. Resources Policy. ISSN 03014207. Vol. 38:1 (2013), p. 36–43. doi: 10.1016/j.resourpol.2012.07.001.
- MATEUS, A. - Recursos naturais de lítio. Revista de Ciência Elementar. ISSN 2183-1270. Vol. 8:3 (2020), p. 1–34. doi: 10.24927/rce2020.034.
- MORAIS, P. M. B. - Os SIG no processo de criação de instrumentos de apoio à decisão. Lisboa: ULisboa, 2013. Master thesis.
- MÜLLER, A.; WANVIK, J. E.; IHLEN, P. M. - Chapter 4 Petrological and Chemical Characterisation of High-Purity Quartz Deposits with Examples from Norway. In *Quartz: Deposits, mineralogy and analytics*. Berlin: Springer-Verlag, 2012. ISBN 9783642221613.
- MÜLLER, A.; ROMER, R. F.; AUGLAND, L. E.; ZHOU, H.; ROSING-SCHOW, NANNA.; SPRATT, J.; HUSDAL, T. - Two-stage regional rare-element pegmatite formation at Tysfjord, Norway: Implications for the timing of Late-Svecofennian and Late-Caledonian high-temperature events. International Journal of Earth Sciences (in review). (2021).
- MÜLLER, A.,[et al.] - NORWEGIAN PEGMATITES I: Tysfjord-Hamarøy, Evje-Iveland, Langesundsfjord. 6th ed. Trondheim : Norsk Geologisk Forening, 2017. ISBN 9788283470208.
- MÜLLER, A.C.; GUIDO, S. - Introduction to machine learning with Python: A guide for data Scientists. Sebastopol: O’ Reilly Media Inc, 2016. ISBN: 9781449369415.

- NEAVE, D. A.; BLACK, M.; TEAL R. R.; GIBSON, S. A.; FERRIER, G.; WALL, F.; BROOM-FENDLEY, S.- On the feasibility of imaging carbonatite-hosted rare earth element deposits using remote sensing. *Economic Geology*. ISSN 15540774. Vol. 111:3 (2016), p. 641–665. doi: 10.2113/econgeo.111.3.641
- NIBIO - Norge i bilder [Em linha] [19 mar 2021]. Available at: <<https://www.norgeibilder.no/>>.
- OLDERSHAW, C. - Gemstones. In *Encyclopedia of Geology*. Oxford: Elsevier, 2005. ISBN 978-0-12-369396-9. p. 6–13. doi: 10.1016/B0-12-369396-9/00240-9
- OLESEN, O.; BRÖNNER, M.; EBBING, J.; GELLEIN, J.; GERNIGON, L.; KOZIEL, J.; LAURITSEN, T.; MYKLEBUST, R.; PASCAL, C.; SAND, M.; SOLHEIM, D.; USOV, S.- New aeromagnetic and gravity compilations from Norway and adjacent areas: methods and applications. *Geological Society, London, Petroleum Geology Conference series*. Vol. 7:1 (2010), p. 559–586. doi: 10.1144/0070559
- PANTELEEV, V. V.; PUTRENKO, O. I.; YANKOVSKII, A. A. - Laser spectral analysis of mineral samples. *Journal of Applied Spectroscopy* . ISSN 1573-8647. Vol. 62:6 (1995), p. 989–995. doi: 10.1007/BF02606745.
- PATEL, C. M.; PATEL, C. D.; RAMI, J. M.; PATEL, K.M.- Optical spectroscopic study of natural rock's minerals. *Materials Today: Proceedings*. ISSN 22147853. Vol. 43 (2020), p. 497–501. doi: 10.1016/j.matpr.2020.12.022.
- PAVLOV, Y. L. - *Random Forests* [Em linha]. Netherlands: De Gruyter, 2020. [7 Apr 2021] Available at:<URL:<https://books.google.pt/books?id=HBmBDwAAQBAJ>>. ISBN 9783110941975.
- PEI, T.; XU, J.; LIU, Y.; HUANG, X.; ZHANG, L.; DONG, W.; QIN, C.; SONG, C.; ZHOU, C.; GONG, J. - Current and future roles of GIScience and remote sensing in natural resource and environmental research. *Geography and Sustainability*. ISSN 26666839. Vol. 2:3 (2021), p. 207-215. doi: 10.1016/j.geosus.2021.08.004

- PEDREGOSA, F.; WEISS, R.; BRUCHER, M. - Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. ISSN 15324435. Vol. 12 (2011), p. 2825–2830.
- PHAN T. M.; KAPPAS, M. - Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors. ISSN 1424-8220. Vol. 18:1 (2018). p. 1-20. doi: 10.3390/s18010018.
- POUR, A. B.; HASHIM, M. - Hydrothermal alteration mapping from Landsat-8 data: Sar Cheshmeh copper mining district, south-eastern Islamic Republic of Iran. Journal of Taibah University for Science. ISSN 1658-3655. Vol. 9:2 (2015), p. 155–166. doi: <https://doi.org/10.1016/j.jtusci.2014.11.008>.
- RAYNER, J. N. - Spectral Analysis. In International Encyclopedia of the Social & Behavioral Sciences. [Em linha]. Oxford : Pergamon, 2001. Available at: <[URL:https://www.sciencedirect.com/science/article/pii/B0080430767025146](https://www.sciencedirect.com/science/article/pii/B0080430767025146)>. ISBN 978-0-08-043076-8. p. 14861–14864
- RICHARDS, J. A.; JIA, X. - Remote sensing digital image analysis: An introduction. [em linha]. 4th ed. Germany: Springer, 2006. [16 Jun 2021] available at: <<https://link.springer.com/book/10.1007/3-540-29711-1>>. ISBN-10 3-540-25128-6. doi: 10.1007/3-540-29711-1.
- RODRÍGUEZ, L.; SEBASTIÁN, J.; TOSI-FURTADO, A.; ALTIMIRAS-MARTIN, A. - Materias primas críticas y complejidad económica en América Latina. Apuntes del Cenes. ISSN 0120-3053. Vol 37:65 (2018), p. 15–51. doi: 10.19053/01203053.v37.n65.2018.5426
- SALLES, R. R.; SOUSA-FILHO C. R.; CUDAHY, T., Vicente, L. E.; MONTEIRO, L. V. S.- Hyperspectral remote sensing applied to uranium exploration: A case study at the Mary Kathleen metamorphic-hydrothermal U-REE deposit, NW, Queensland, Australia. Journal of Geochemical Exploration. ISSN 03756742. Vol. 179 (2017), p. 36–50. doi: 10.1016/j.gexplo.2016.07.002

- SANTOS, D.; TEODORO, A. C. M.; CARDOSO-FERNANDES, J.; LIMA, A.; - Remote sensing techniques to detect areas with potential for lithium exploration in Minas Gerais, Brazil. *Proc. SPIE. Earth Resources and Environmental Remote Sensing/GIS Applications X*. ISSN 1996756X. Vol. 1:1 (2019), p. 50. doi: 10.1117/12.2532744
- SIMON, N.; Ali, C.; Roslan, K.; Sharir, K. - Best Band Ratio Combinations for the Lithological Discrimination of the Dayang Bunting and Tuba Islands, Langkawi, Malaysia. *Sains Malayasiana*. ISSN 15325415. Vol. 45:5 (2016), p. 659–667.
- SINGH, A.; HARRISON, A. - Standardized principal components. *International Journal of Remote Sensing*. ISSN 13665901. Vol. 6:6 (1985), p. 883–896. doi: 10.1080/01431168508948511
- SOUZA, R. B. - Sensoriamento Remoto : conceitos fundamentais e plataformas [Em Linha]. Santa Maria: Instituto Nacional de Pesquisas Espaciais(INPE), 2010. [5 Jun 2021] Available at <http://www3.inpe.br/crs/crectalc/pdf/ronald_ceos.pdf>.
- TEODORO, A. C.; SANTOS, D.; CARDOSO-FERNANDES, J.; LIMA, A.; BRÖNNER, M. - Identification of pegmatite bodies, at a province scale, using machine learning algorithms: preliminary results. *Proc. SPIE - Earth Resources and Environmental Remote Sensing/GIS Applications XII*. ISSN 1996756X. Vol. 1186308 (2021), p. 6. doi: 10.1117/12.2599600.
- TOTH, C.; JÓZKÓW, G. - Remote sensing platforms and sensors: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*. ISSN 09242716. Vol. 115 (2016), p. 22–36. doi: 10.1016/j.isprsjprs.2015.10.004
- VACARI, A. L.; SANTOS, L. C.; FRÓES, N. J.; NASCIMENTO, D. N.- OS DESAFIOS PARA A IMPLANTAÇÃO E MANUTENÇÃO DE ENERGIAS VERDES : UM PANORAMA INTERNACIONAL. [Em Linha]. Bauru: XXV Simpósio de Engenharia de Produção (SIMPEP), 2018. [13 Jun 2021]. available at: <https://www.researchgate.net/publication/328930137_Os_desafios_para_a_implantacao_e_manutencao_de_energias_verdes_um_panorama_internacional>

VAPNIK, V. - The nature of statistical learning theory. 2nd ed. New York : Springer science
& business media, 2000. ISBN: 9781475724424

YU, L.; PORWAI, A., HOLDEN, E. J.; DENTITH, M. - Towards Automatic Lithological
Classification from Remote Sensing Data Using Support Vector Machines.
Computers & Geosciences journal. ISSN 0098-3004. Vol. 45 (2012), p. 229–239. doi:
10.1016/j.cageo.2011.11.019

ZIMMERMANN, R.; Melanie B.; Louis A.; Kombada M.; Richard G. - Remote sensing
exploration of Nb-Ta-LREE-enriched carbonatite (Epembe/Namibia). Remote
Sensing. ISSN 20724292. Vol. 8:8 (2016), p. 620. doi: 10.3390/rs8080620

Attachment

PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

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

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

Ana Claudia M. Teodoro, Douglas Santos, Joana Cardoso-Fernandes, Alexandre Lima, Marco Brönnner, "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}, Brönnner, M.³

¹Institute of Earth Sciences, FCUP pole, Porto, Portugal; ²Department of Geosciences, Environment and Spatial Planning Faculty of Sciences of the University of Porto, Portugal; ³Geological 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

Proc. of SPIE Vol. 11863 1186308-1

Downloaded From: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie> on 15 Sep 2021
Terms of Use: <https://www.spiedigitallibrary.org/terms-of-use>

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 fluorite⁵. 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^{5,6}.

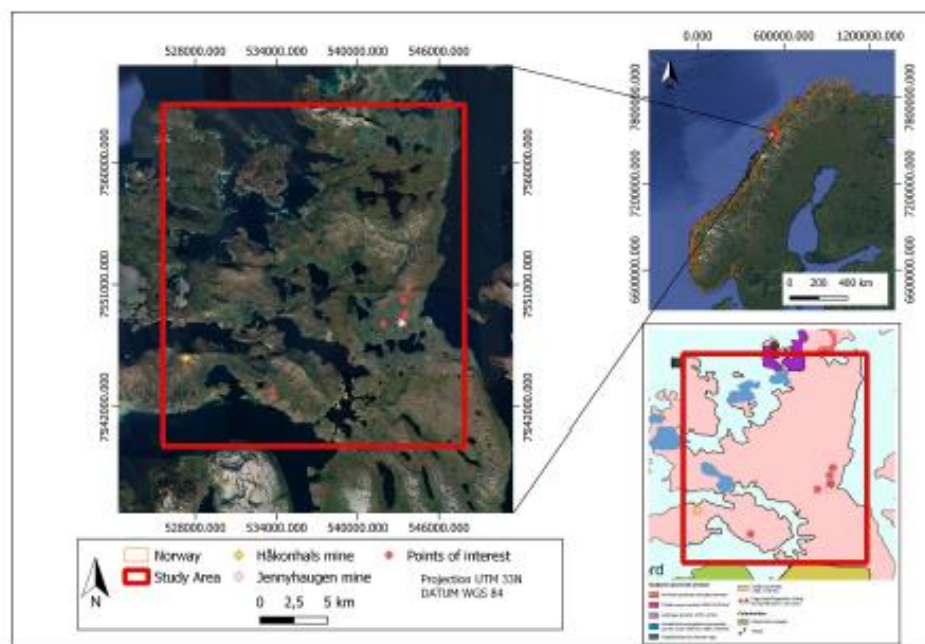


Figure 1. Study area location in Tysfjord district, Norway. The aerial photograph to the left shows the jagged landscape with high topography, intermingled by a large number of waterbodies from lakes and fjords. In the lower right corner, an adapted geological map of the district marks the most prominent 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-learn⁷. 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 accuracy¹¹. 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 input¹². 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 construction¹³. 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 Vapnik¹⁶, and based on a statistical learning theory^{17,18}. 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 Kappas¹⁴ 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 maps⁵, aeromagnetic data¹⁸, 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 (DOSI)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, Spain²¹ and in southeastern Brazil²² 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) PCA²⁵ (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 maps⁵, aeromagnetic data¹⁸, high-resolution aerial photographs²⁰, 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, Eastman³⁰, 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 50¹⁰.

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 class¹⁰.

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.

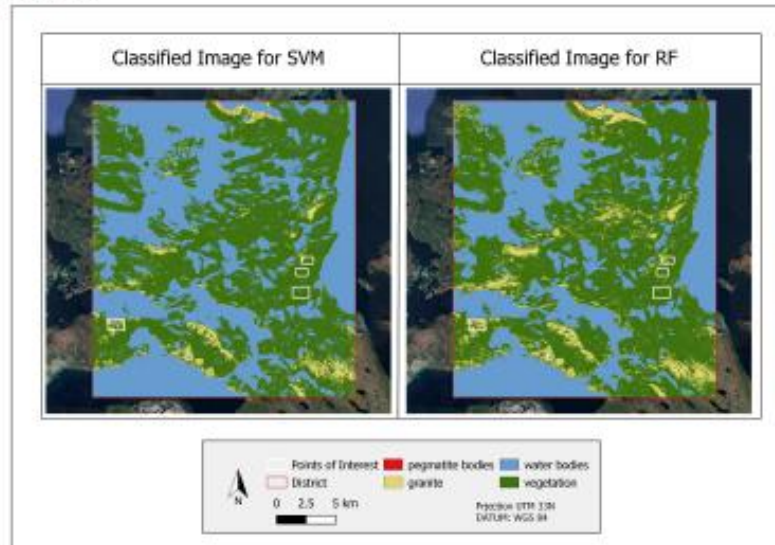


Figure 2. Comparison of results for the SVM (left) and RF (right) algorithm classifications.

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.

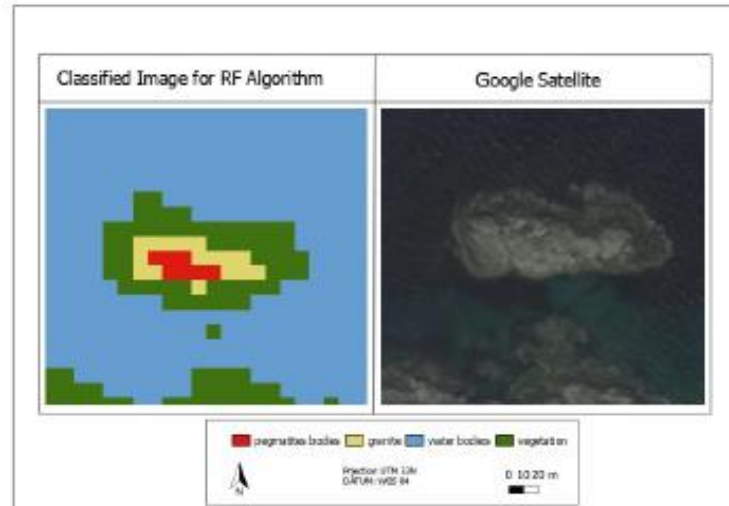


Figure 3. Comparison between RF classification and Google satellite image where RF assumes as vegetation pixels that correspond to water.

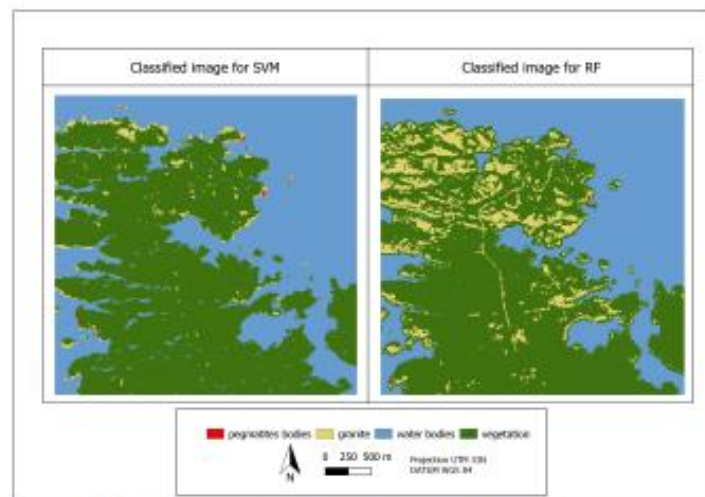


Figure 4. Exemplification of the difference in granite classification between the two algorithms.

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.

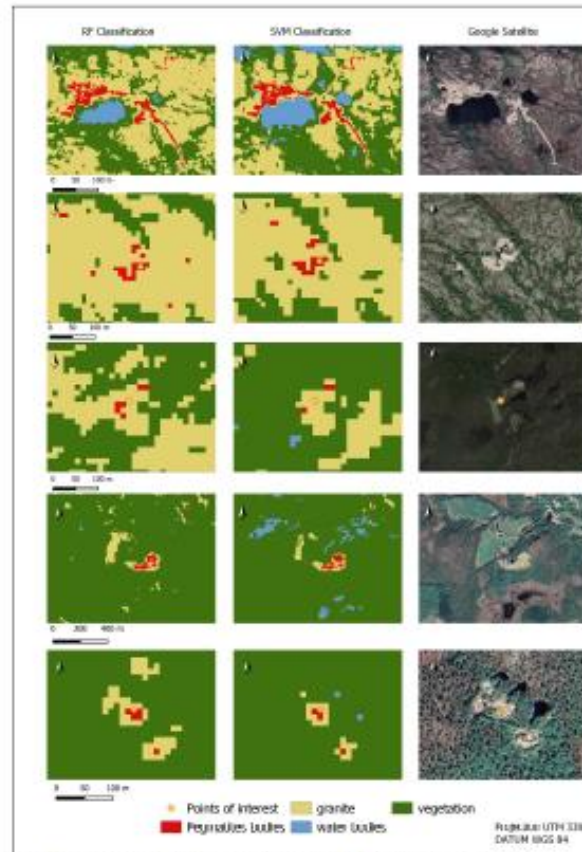


Figure 5. Comparison of the classification of the two algorithms for the known pegmatite areas.

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.

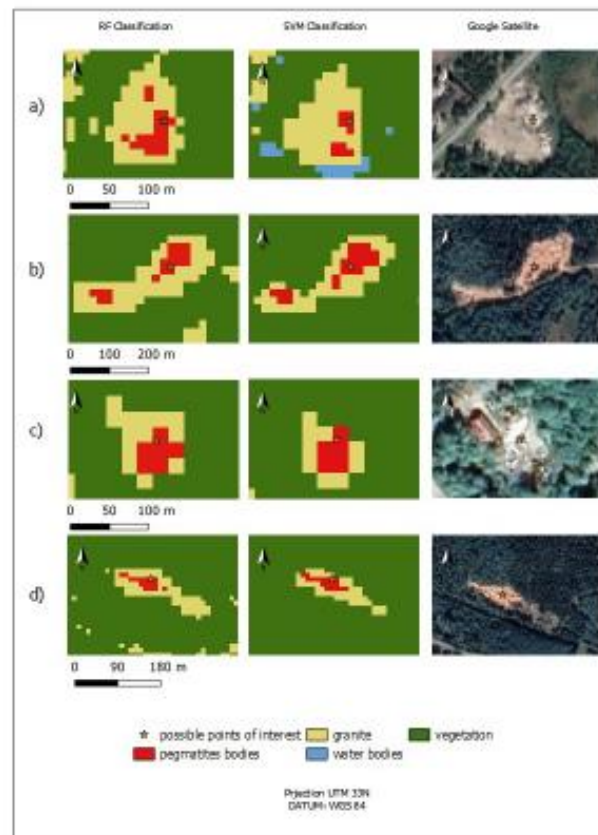


Figure 6. Comparison of the prominence of possible areas with the presence of pegmatites.

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.

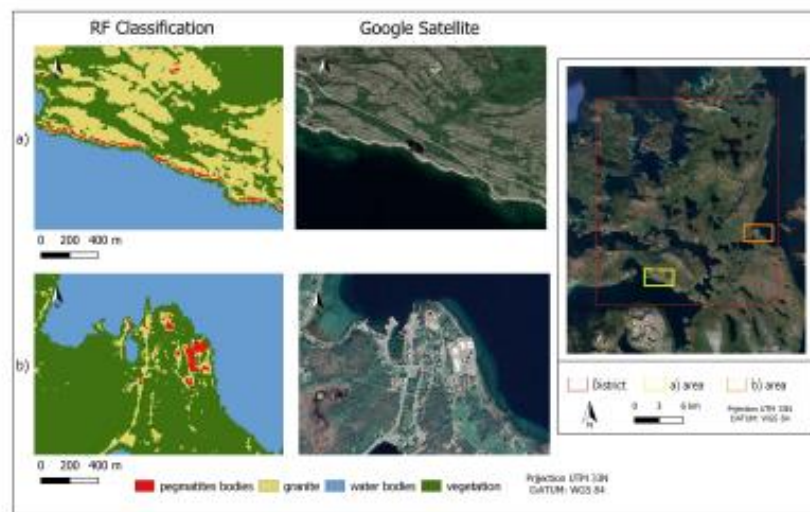


Figure 8. Examples of false positives obtained with the RF algorithm: a) an example of the signal confusion in a coastal area; b) an example of the signal confusion between buildings and pegmatites.

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 assessment], Volume I, 387 (2010).
- [5] Hamarøy, T., Iveland, E., Müller, B. A., Husdal, T., Sunde, Ø., Fris, H., Andersen, T., Johansen, T. S., Werner, R., Thoresen, Ø. and Olerud, S., [Norwegian Pegmatites I: Tysfjord-Hamarøy, Evje-Iveland, Langesundsfjord], Geol. Society Norw., Trondheim, 122 p. (2017).
- [6] Husdal, T., Müller, A., Olerud, S. and Thoresen, Ø., "1. Pegmatites of the Tysfjord-Hamarøy area, northern Norway", [Norwegian Pegmatites I: Tysfjord-Hamarøy, Evje-Iveland, Langesundsfjord], Husås, A. M., Ed., Geol. Society 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., Brø 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] Loughlin 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.norgebilder.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).