Masters
Program
in Geospatial
Technologies

MONITORING AND EVALUATION OF THE BEST MAPPING APPROACH OF THE ENVIRONMENTAL IMPACTS AND LAND COVER CHANGES CAUSED WITH THE REVEAL OF THE QUARRY USING THE LANDSAT AND UAV IMAGES

Nikolina Nišić

Dissertation submitted in partial fulfilment of the requirements
for the Degree of *Master of Science in Geospatial Technologies*

MONITORING AND EVALUATION OF THE BEST MAPPING APPROACH OF THE ENVIRONMENTAL IMPACTS AND LAND COVER CHANGES CAUSED WITH THE REVEAL OF THE QUARRY USING THE LANDSAT AND UAV IMAGES

Dissertation supervised by

Ph.D. Roberto Henriques, Ph.D. Edzer Pebesma and Ph.D. Joaquin Torres Sospedra

February 2018

ACKNOWLEDGMENTS

I would first like to thank my thesis supervisors Ph.D. Roberto Henriques (NOVA IMS, Lisbon), Ph.D. Edzer Pebesma (Institute for Geoinformatics, Münster) and Ph.D. Joaquin Torres Sospedra (Universitat Jaime I, Castellon), for support and guidance throughout the whole period of the master thesis. The door to Prof. Henriques office was always open whenever I ran into a trouble spot or had a question about my research or writing. He allowed this paper to be my own work but, steered me in the right the direction whenever he thought I needed it.

I would also like to thank my professor Ph.D. Miro Govedarica (Faculty of Technical Sciences, Novi Sad) who provided me data, that were used for the execution of this project. Without his help and support the practical execution of this project could not have been successfully conducted.

I would also like to acknowledge Ph.D. Marco Painho of the NOVA IMS School at New University of Lisbon as the second reader of this thesis, and I am gratefully indebted for his very valuable comments and help on this thesis.

Finally, I must express my very profound gratitude to my family and friends for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Author,

Nikolina Nišić

MONITORING AND EVALUATION OF THE BEST MAPPING APPROACH OF THE ENVIRONMENTAL IMPACTS AND LAND COVER CHANGES CAUSED WITH THE REVEAL OF THE QUARRY USING THE LANDSAT AND UAV IMAGES

ABSTRACT

The major incidents caused from mine waste and poor monitoring of the quarry sites in Serbia has caused the formation of a large number of contaminated areas that are causing environmental and social impacts and land cover changes. Related impacts in the European countries has initiated the formation of the environmental legislation for the prevention of the quarry risks on the environment. This project evaluates the EU MWD Protocol on the real case of the quarry site "Gradinje". The protocol is linked with the GIS system and showed the presence of the contamination, which later was mapped using the most adequate classification approach. Pixel based image analysis approaches for classifying land cover classes are compared using three supervised classification algorithms: maximum likelihood (ML), support vector machine (SVM) and random forest (RF). Concerning the comparison between different classification predictions, the McNemar test indicated that the observed difference between the two machine learning algorithms (RF and SVM) was not statistically significant ($p = 0.2278 > 0.05$). However, there was not a statistical significant difference ($p > 0.05$) in classification accuracy between map produced using SVM algorithm and RF algorithm. Classification based on SVM and RF algorithms produced map that is more visually adequate depiction of land cover classes than that produced by ML algorithm. Overall, two machine learning algorithms seems superior than ML algorithm in land cover classification using medium spatial resolution imagery in term of performance based on classification accuracy.

Praćenje uticaja kamenoloma na životnu sredinu i društvo korišćenjem različitih metoda mapiranja terena, upotrebom Landsat i slika bespilotnih letjelica

SAŽETAK

Značajni incidenti prouzrokovani minskim otpadom i lošim praćenjem nalazišta kamenoloma u Srbiji doveli su do stvaranja velikog broja zagađenih površina koje izazivaju ekološke i društvene uticaje i promjene zemljišnog pokrivača. Srodni uticaji u evropskim zemljama pokrenuli su formiranje ekološkog zakonodavstva za sprečavanje rizika kamenoloma na životnu sredinu. Ovaj master rad vrši primjenu unapred definisanog protokola o rudnom otpadu na stvarnom slučaju kamenoloma "Gradinje". Protokol je povezan sa GIS sistemom i pokazao je prisustvo zagađenosti, koja je naredno mapirana koristeći najadekvatniji algoritam klasifikacije. Poređenje pikselskih pristupa za klasifikaciju promjena zemljišnog pokrivača izvršeno je koristeći tri izabrana algoritma: metodu maksimalne vjerodostojnosti (ML), metodu pridruženih vektora (SVM) i metodu nasumične šume (RF). Što se tiče poređenja između različitih rezultata klasifikacije, McNemarov test je pokazao da posmatrana razlika između dva mašinska algoritma (RF i SVM) nije bila statistički značajna ($p = 0,2278 > 0,05$). Međutim, statistički značajna razlika ($p > 0.05$) nije pronađena u tačnosti klasifikacije između mapa proizvedenih korišćenjem SVM algoritma i RF algoritma. Klasifikacija zasnovana na SVM i RF algoritmima je formirala mapu koja vizuelno nudi adekvatno prikazivanje različitih klasa zemljišnog pokrivača, te koja se veoma razlikuje od one proizvedene algoritmom ML. Sve u svemu, dva algoritma koji pripadaju mašinskim algoritmima izgledaju superiorniji od algoritma ML u klasifikaciji zemljišnog pokrivača koristeći slike srednje prostorne rezolucije u smislu performansi zasnovane na preciznosti klasifikacije.

KEYWORDS

Environmental risk

Geographical Information Systems

Landsat Imagery

Maximum Likelihood

Quarry risk assessment

Random Forest

Remote Sensing

Support Vector Machine

The EU MWD Pre-Selection Protocol

Unmanned Aerial Vehicle

LISTA SKRAĆENICA

Bespilotna letjelica

Daljinska detekcija

Ekološki rizik

Evropska direktiva o rudarskom otpadu

Geoinformacioni sistemi

Metoda maksimalne vjerodostojnosti

Metoda nasumične šume

Metoda potpornih vektora

Procjena rizika od kamenoloma

Snimci Landsat satelita

ACRONYMS

AUC- Area Under Curve

DN- Digital Number

DSM - Digital Surface Model

DTM- Digital Terrain Model

ETM+ - Enhanced Thematic Mapper Plus

EU MWD- European Mine and Waste Directive

GPS – Global Positioning System

ISODATA - Iterative Self Organizing Data Analysis Technique

ML- Maximum Likelihood

NDVI- Normalized Difference Vegetation Index

RF- Random Forest

ROC - Receiver Operating Characteristics

SAVI- Soil Adjusted Vegetation Index

SVM- Support Vector Machines

TIRS - Thermal InfraRed Sensor

TSAVI - Transformed Soil Adjusted Vegetation Index

UAV- Unmanned Aerial Vehicle

UTM – Universal Transverse Mercator

WGS - World Geodetic System

INDEX OF THE TEXT

INDEX OF TABLES

INDEX OF FIGURES

1. INTRODUCTION

The main environmental effects caused with the discovery of the quarry are affecting the agricultural land by modification and pollution of the nearby, pre-existing ecosystems. With regards to the prevailing environmental legislation and its enforcement, there is total lack of efforts in monitoring, rehabilitation, restoration or post-mining programmes for minimization of adverse environmental impacts (Lad R. J. and Samant J. S. 2014). Traditional ways of monitoring the quarry using the different surveying data acquisition techniques is very time consuming and cost effective. Remote Sensing data and tools have been used lately as a new cost-effective technique for mapping and monitoring of the land cover changes. The advantage of using this technique is representation of the large areas at high level of detail on spectral and spatial level, thus it made easier and increased the interest in the monitoring of the environmental effects and risks of the open pit-mine on the surrounding area. Most of the negative environmental impacts are created when the quarrying is not performed according to the official legislation. The major environmental and socio-economic problems related to quarrying during this study include, landscape alteration, hill cutting affecting local biodiversity, generation of unproductive wastelands, dust pollution, noise pollution, illegal stone extraction, accidents and in some areas lowering of groundwater table (Maponga and Munyanduri, 1998). The reveal of the quarry affects the environment through the process of quarrying which involves excavation and blasting, transport and disposal of mine waste.

For successful temporal mapping of the land cover changes, it is important to select the appropriate classification technique. Various classification algorithms sub pixel, object based, machine learning or traditional algorithms were used for mapping of the open pit mine. Previously, classical probabilistic based algorithms (eg: maximum likelihood) were used abundantly; but the applicability of these algorithm is not reliable due to its nature to deal only on normally distributed data (Belgiu & Drăgu, 2016). Nonparametric machine learning algorithms is gaining popularity for classification purposes, as they are not relied on arbitrary data distribution (Immitzer et al., 2012). Among others, random forest and support vector machine are considered as widely used algorithms for classifying highly dimensional data (Li et al., 2013).

1.1 Motivation

The motivation for conducting this type of project is relied on several reasons. Firstly, environmental effects caused with the quarrying activities have lack of effort in the monitoring, post-mining activities, risk assessment and evaluation of the mining activities on the surrounding area. Rapid growth of the construction activity in order to meet the modern requirements of construction of the highways and buildings in surrounding cities has enormously increased demand for building materials in area of city Dimitrovgrad. For the afore mentioned purposes, building material is mainly taken from the quarries. On the other hand it creates changes on the environmental and social problems when quarry becomes revealed (Lad R. J. and Samant J. S. 2014). Major environmental effects are destruction of vegetation, disruption of animal habitats, diversion and blockage of natural drainage systems, soil erosion and river siltation, noise and vibration; and dust pollution (Maponga and Munyanduri, 1998). Existence of the afore mentioned environmental effects has led to the conflicts between the mine owners and the locals living near the quarry sites. The issue gained media attention but unfortunately no actions have been taken regarding to that, therefore my motivation has arisen to perform the quarry reveal analysis and to evaluate the environmental risks of the quarry and its impact on the population and to put the analysis into action in order to prevent in future the similar actions and results. According to the interviews of the inhabitants the excavating activities and explosions in the quarries were strong enough to cause the damage of the nearby houses and as well as to increase the effects of the environmental noise on the nearby inhabitants (Figure 1).

Figure 1. Damages on the houses caused by quarry excavation activities, Resource:http://rs.n1info.com/a210245/Vesti/Vesti/Dimitrovgrad-Mestani-cekaju-odstetu-Koridora.html

Secondly, the classification results may provide sufficient and reliable information for developing the quarry monitoring processes, which can be proved substantial on prevention of the negative effects of the quarrying on the surrounding area. Finally, the outcome of the project could be a significant source of information for other purposes (research etc.), estimating the current situation of the land cover changes with the quarry reveal as well as the evaluation of environmental effects of the quarry in spatial terms through time and the potential essential rehabilitation of some specific but critical areas (e.g. natural habitats etc.) for minimization of the environmental effects of the quarry.

It should be highlighted the fact that even though there are few studies about the comparative effectiveness of the different classifying algorithms, most of them are focused on purely land cover classification. Thus, it is observed that there is a relatively research gap in the classification of the open pit mines. On the other hand, the benefits of this type of project far exceeds its disadvantages providing new and constructive solutions to society protecting the valuable resources of the natural environment. To this end, robust classifiers are used in a comparative manner, so that we may explore the effectiveness of them and adopt the most suitable one in similar projects. The added value of this project is reflected on the flexibility of applicability to any study area after the necessary adjustment to local peculiarities.

1.2 Goals and objectives

The aim of the thesis is to find out which one of the three selected machine learning algorithms gives the most accurate results in the context of land cover changes caused with the impact of the quarry reveal. The comparison of the following algorithms: maximum likelihood (ML approach) and machine learning algorithms [i.e. support vector machine (SVM) and random forest (RF)] was performed in order to classify the study area using satellite imagery with spatial resolution of 30 m. Visual inspection and statistical assessment of classification outputs; and performance evaluation of the classification models are basic guidelines for comparison. The main evaluation parameters used to compare afore mentioned classifiers, are overall accuracy, kappa coefficient, MC Nemar test and the time of execution of each classifier. The experimentation and execution of the

main goal of this thesis is done using the following packages of R software: 'randomForest' package for the execution of the RF algorithm while for the SVM classification 'e1071' package was retrieved. All classification models were developed using the 'caret' packages (Kuhn, 2016) for single consistent environment for training and parameters tuning. The input data for the software were Landsat images, and output are classified images of the land cover for the selected study area.

Sub objective of this project is to evaluate the impact of the quarrying activities and quarry itself on the surrounding environment. For that purpose, tasks were to organize, test and evaluate RGB data collection using the UAV that are suitable for the modelling of the quarry. Creation of the DSM, orthophoto and point cloud mesh is used for the achieving the clearer insight in study area and it characteristics. Basic analysis of the study area were perfomed in order to meet the requirements and to perform the analysis of risk assessment according to the local thresholds of EU MWD (Mine Waste Directive) pre-selection Protocol.

1.3 Study Area

Considering the need to cover different types of land cover classes and changes in it caused with the open pit mine excavations in order to test my mapping approach the study area in the southeastern part of Serbia in the district Pirot was chosen (Figure 2.). The study area is distanced about 2km from the border with Bulgaria. It covers about 1000 square meters. The prevalent land cover types of the study region primarily constitute of mixed forest types, cultivation fields, water bodies, human settlement and quarry sites. In the study area live 178 adult people, with the average age of the population as follows 47.2% for the male and 47.3% for the female residents. There are 84 households and the average number of the people in the household is 2.43. Population in this area is very heterogeneous, with majority of Bulgarian population since the settlement is located very close to the border. According to the censuses from the previous years it is slightly in decrease.

The quarry events affected some types of land cover accompanied with multiple environmental and socioeconomic consequences with significant changes at two locations in the central and northern part of the study area. The presence of two quarry sites within the short distances to the nearby settlements and the water source, makes reasonable the exploration and monitoring of the impacts and potential risks of the quarry activities.
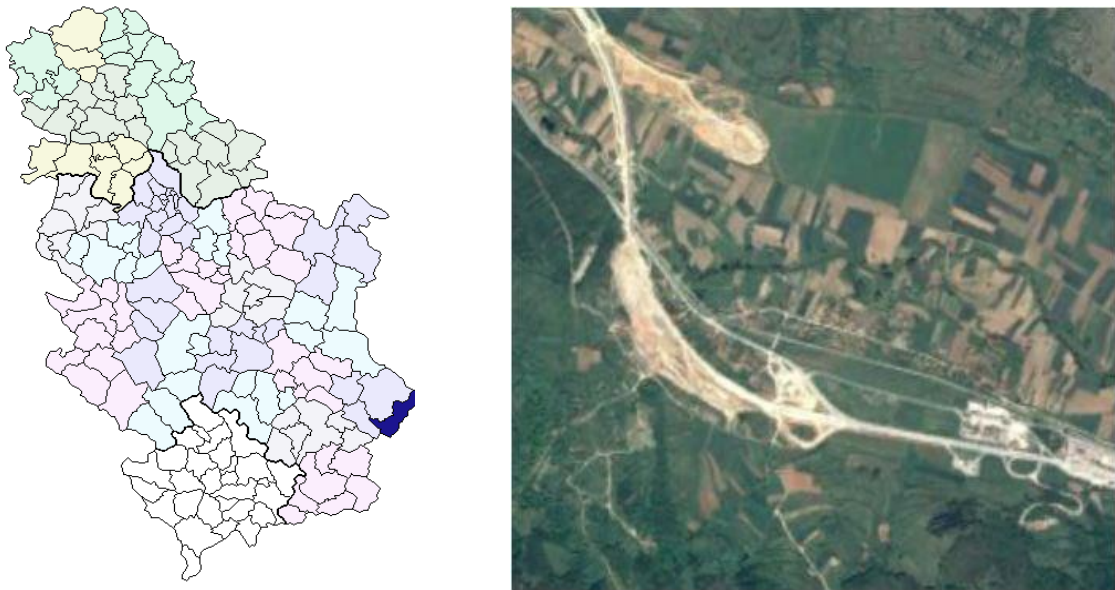
Figure 2. Study area, Author: Nisic 2018, Resources: https://sr.wikipedia.org/

2. ENVIRONMENTAL EFFECTS AND POSSIBLE RISKS OF THE ACTIVITIES IN QUARRY SITE

2.1 Theoretical framework

The most obvious engineering impact of quarrying is change in geomorphology and conversion of land use, with the associated change in visual scene (Langer 2001). Most engineering impacts can be controlled, mitigated, kept at tolerable levels and restricted to the immediate vicinity of the aggregate operation by employing responsible operational practices that use available engineering techniques and technology (Langer 2001). Residents near quarry sites have been interviewed in order to get their opinion and views regarding the quarry operations in Vellarada panchayat and according to the results the shortage of water and pollution were the main negative impacts of quarrying (Chandran 2015). As well as during the rainy season the wells are flooded, and the water level remained in a stagnant position up to several months but after quarry function started even in rainy month also water level goes down rapidly. Maximum of the respondents complained about the problems related to skin and respiratory diseases (Chandran 2015). According to the residents responds in the final conclusion majority of them were against quarry operation close to inhabited areas.

Surface mining activities range from large open-cast coal and base metal mines, too much smaller aggregate (rock, gravel and sand), industrial minerals (potash, clay) and building materials (granite, stone and marble) quarries. Mapping mining activities and evaluating associated environmental concerns are difficult problems because of the extensive area affected and the large size of individual mines. Monitoring and controlling these changes have been more difficult because of the expense and time in producing reliable and up-to-date mapping (Saroglu et al. 2005). The purpose of digital land cover classification is to link the spectral characteristics of the image to a meaningful information class value, which can be displayed as a map so that resource managers or scientists can evaluate the landscape in an accurate and cost-effective manner (Weber and Dunno 2001).

For the purpose of detecting the land cover changes caused with the discovery of the quarry in the article of Saroglu et al. (2005), unsupervised classification technique ISODATA (Iterative Self Organizing Data Analysis Technique) was performed, taking into account 4 land cover types: forest, water, stone quarry and others. Among the afore mentioned land cover types the drastically change was the loss of the forest and significant increase of the stone quarry sites over the period of 14 years. Although mining activities should be organized by terms of sustainable development, mining activities have been executed illegally. In this concept, the distribution and expansion of these activities should be monitored, their environmental damages should be determined; mitigation studies against these damages should be performed, and regular inspections should be executed to keep these activities under control (Saroglu et al. 2005).

Line drilling and sawing are modern techniques for quarrying. Line drilling (also called slot drilling) consists of drilling a series of overlapping holes using a drill that is mounted on a quarry bar or frame that aligns the holes and holds the drill in position. Sawing can be accomplished with a variety of saws including wire saws, belt saws, and chain (Langer 2001).

Nowadays, the valuation of the impact on the environment represents an integral part of the mining activities and, especially, of the surface ones (Latifovic et al., 2005).

Remote sensing data allow for the identification and monitoring of pollution sources (Ress and Williams, 1997) and affected surfaces including abandoned land and the changes in using the land and the water courses (Schmidt and Glaesser, 1998). Remote sensing, by synoptic cover on synchronized intervals, is very useful in monitoring the perturbations caused by mining activities (Irons et al., 1980; Parks et al., 1987; Wier et al., 1973), reforesting surfaces exploitations and monitoring regenerated lands (Kandrika and Dwivedi, 2003; Legg, 1986).

In the article of the Vorovencii (2011) it has been analyzed the effects that limestone exploitations in the outskirts of Braşov city have on the landscape, using RGB combination, vegetation indices and spectrum profiles. A series of Landsat 5 TM satellite images taken in 1984, 1989 and 2009 has been used. The results obtained show that the most adequate RGB combination for limestone quarries identification on this type of

images is the combination that includes NDVI (Vorovencii 2011). In the analysis of Vorovencii (2011) of the impact of the limestone quarries on the environment, it has suggested   that different NDVI indices such as TSAVI (Transformed Soil Adjusted Vegetation Index) and SAVI (Soil Adjusted Vegetation Index) can help with the distinction of the concreate and asphalt in the nearby places from the limestone quarry which was the subject of monitoring and mapping. Detailed studies of limestone quarry regression analysis together with the spectrum profiles has showed that limestone quarries are having more significant reflectance in the red band as compared with the near infrared band.

 The consequences of deforestation on the Plateau region due to mining are diverse with various environmental and economic implications. It has led to the loss of forest products, flooding and erosion, aggravation of desertification and pollution, high rate of siltation of streams, decline of biological diversity and acceleration of extinction of the Plateau region flora and Fauna (Oyinloye & Ajayi 2015).

Based on the article of Oyinloye & Ajayi (2015) regarding the research of the environmental impacts of quarry on residents it has been proved that the residents living around 1km from the quarry were very exposed to effects of the pollution, cracking of the building noise as well as the presence of the some health challenges, where 24.8% of the residents living in the buffer zone of 3km (vulnerable area according to the NEA Singapore) from the quarry have reported prevalence throat infections, 39% of them had cough/catarrh, 14.2% reported the issue of the heart diseases, 17.7% experienced chest aches and 4.3% reported the existence of skin infections. According to the recommendations of this research the government should act in accordance with the pollution control policy and forbid quarrying activities in buffer zone of 3km in order to prevent the rate of inhalation of harmful substances by the residents.

In the study of Koruyan et al. (2012), changes in the natural vegetation as a function of production rate were determined using the NDVI. To determine the extent of vegetation loss, NDVI images from one year to the next were subtracted from the AOI images. According to the author only few studies have been performed in order to assess the environmental risks of the quarry using the remote sensing.

Hazard identification and risk assessment of the quarrying is a continual process. The main causes of the quarrying hazard might occur from the following mining operations such as falling of sides, drilling operation, blasting operation, transportation and some of the natural hazards (Towl 2005).

2.2 The EU MWD Pre-selection Protocol

Major changes in the land cover as well as increase of the contaminated mine sites influenced the development of the EU environmental legislation in order to minimize potential risks of such incidents. EU commission has developed the legislation named EU MWD (Mine Waste Directive) pre-selection Protocol which consist of the 18 simple questions about contamination source, pathways and receptors. Questions of EU legislation have been incorporated to the GIS system, together with the crucial elements a such as the topographic slope and distance to the nearest surface and groundwater bodies to settlements and protected areas are calculated and statistically evaluated in order to adjust the Risk Assessment models to local conditions in Hungary (A.Abdal et al 2012).

The European Union Directive on the Management of Waste from the Extractive Industries [Mining Waste Directive (MWD); Directive 2006/21/EC] requires in Article 20 that "Member States shall ensure that an inventory of closed waste facilities, including abandoned waste facilities, which cause serious negative environmental impacts or have the potential of becoming in the medium or short term a serious threat to human health or the environment is drawn up and periodically updated." According to Article 21, such methodologies shall allow for the establishment of the most appropriate risk assessment procedures and remedial actions having regard to the variation of geological, hydrogeological and climatological characteristics across Europe.

Contamination RA is defined as the probability of adverse effects to humans and ecosystem resulting from exposure to environmental pollutants (Kolluru et al. 1996; Fergusson 1998; US EPA 1989, 1998; Di Sante et al. 2009; Fan et al. 2010) RA includes the steps of (1) hazard description, (2) dose/response (toxicity) analysis, (3) contaminant transport, (4) exposure assessment, (5) risk characterization and (6) risk management (Van

Leuwen and Hermens 1996; U.S. EPA 2002, 2007). The Protocol thresholds are based on the Irish regulation for the operation of ponds with respect to quarries (Safe Quarry 2008).

The EU MWD Pre-selection Protocol (Stanley et al. 2011) is based on a 'YES-or-NO' questionnaire (Appendix) and consists of 18 questions using simple criteria available in existing databases readily enabling the preliminary screening of the mine waste sites for environmental risk. This questionnaire is used to eliminate the quarry sites which are not dangerous for the environment and population. The questions from the afore mentioned Protocol can be answered with 'YES', 'NO' or in presence of uncertainty 'UNKNOWN'. In case that question is answered with 'UNKNOWN' it would follow the same path as 'YES' answer since there is possibility of contamination and it should lead toward the further examination. In case that we have answer 'NO' for the all question in one section, there is no further need for the examination.

The EU MWD Pre-selection Protocol, as shown in its flowchart (Appendix 1), consists of four sections: (1) Known serious impact, (2) Source, (3) Pathways and (4) Receptors.

2.2.1   First section of the EU MWD Pre-Selection Protocol

First section seeks to explore if there was any accident regarding the health issues in the area around and in the quarry site. It consists of the main question for the all further analysis, as following:

Q1: Is the mine waste facility known to have an incident which had a serious impact on human health or environment (Stanley et al. 2011)?

The first question of the EU MWD Pre-selection Protocol coincides with the base question of my master thesis, and one of the aims why I have chosen this topic and research area. Since the revelation of the quarry "Gradinje" has gained media attention in Serbia but in other Balkan countries, it attracted my attention as well and created a desire to explore more about the situation by combining the free data from the internet such as Landsat images, Google Earth images, polygons of the area and land cover data and the data obtained from the measurements on the field (UAV images). According to the information from medias and the sill images from the field, there is evidence of potential risks of the

quarry on the environment, as well as the presence of the physical damage of the nearby houses caused in opinion of the inhabitants from quarry blasts and excavations. According to the interviews of the inhabitants, residents who are living within the circle of 1km from the quarry site were exposed to the effects of the pollution, so they experienced the several health issues such as throat, cough and skin infections. There was one event of the heart disease of the inhabitant who claimed that there was absence of the pre-existing conditions of such illness before the excavations (http://rs.n1info.com/a210245/Vesti/Vesti/Dimitrovgrad-Mestani-cekaju-odstetu-Koridora.html). Since the data were not confirmed with the medical documentation of the inhabitant, the following examinations of the Protocol were performed.

2.2.2   *Section 2 of the EU MWD Pre-Selection Protocol*

Section 2 seeks to explore the contamination of the quarry site taking into the account chemical and physical structure of the quarry site (source questions Q2-Q10). It is divided in two parts, where the first one tends to give an answer regarding the content of the source pollution while the second focuses on the stability of the quarry. List of the questions that are involved in this section are represented in the Table 2. Since the data about the chemical elements of the quarry structure were not available in the exploration part responses regarding the 2A section were entered as a 'UNKNOWN', which indirectly leads the process to the further examination.

Regarding the section 2B, exploration of the quarry stability was performed using the available UAV images in form of ortho-mosaic and 3D model of terrain. In this section there were performed two analyses that involved creation of the 3D model from the sill UAV images and the simple measurements of the distances between quarry and facility, height of the waste heap, total area of the quarry and the slope of the foundation. The afore mentioned analysis is shown on the following map depicted in Figure 3.
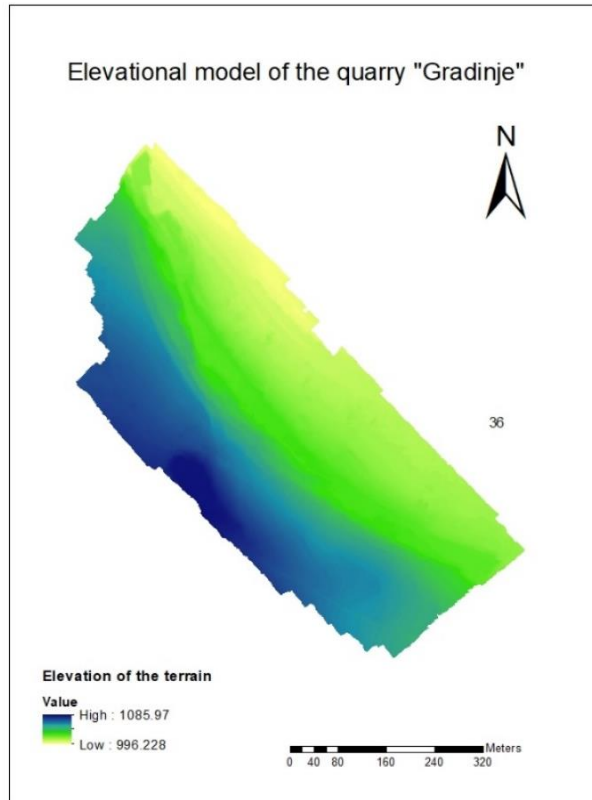
Figure 3. Elevational model of the quarry "Gradinje"

It was not objective of the report only to take into the account the existing information about the quarry site but to perform the analysis of the terrain based on the real data obtained on the field.

Question	Issue	YES	NO	UNKNOWN
2A. SOURCE-contents potential source of			**Is the mine waste facility a pollutants?**	
Q2	Did the mine work sulphide minerals or produce a waste containing sulphide minerals?			UNKNOWN

Q3	Were any of the following produced from the mined mineral-Ag, As, Ba, Be, Cd, Co, Cr, Cu, Hg, Ni, Pb, Sb, Se, Sn, Te, TL, U, V, Zn or asbestos?			UNKNOWN
Q4	Did the mine use dangerous chemicals to process the mined minerals?			UNKNOWN
2B. SOURCE-stability		**Is the source physically stable?**		
Q5	Is the waste facility a tailings lagoon?	YES		
Q6	Is the area of the tailings lagoon > 10 000m2?		NO	
Q7	Is the height of the tailings lagoon > 4m within 50m of the facility?	YES		
Q8	Is the area of the waste heap >10 000m2		NO	
Q9	Is the height of the waste heap >20m?	YES		
Q10	Is the slope of the foundation >1:12?	YES		

Table 1. Questions of the second section of the EU MWD Pre-selection Protocol (Q2 to Q10)

2.2.3   Third section of the EU MWD Pre-Selection Protocol

Third section consist of four questions that are tending to explain the possible pathways from which receptors can have negative impact of the mine waste. In order to get responses in the section 3, there were used 2 types of data: Google Earth images as well as elevational model made from the UAV images.

From the planar measurements on the Landsat images it has been concluded that the surface water body (river) is distanced within the circle of 1km from the quarry sites. However, there is no indication of the groundwater beneath the mine waste, thus this question was marked as 'UNKNOWN'. From the sill images of the UAV, there could be

noticed that the quarry site is uncovered and on significantly higher altitude, making the surrounding agriculture fields and settlement directly exposed to the air transmission of the particles of the quarry material.

Question	Issue	YES	NO	UNKNOWN
3.PATHWAY Are there potential **pathways** for possible contamination to reach potential receptors?				
Q11	SURFACE WATER Is there a water course within 1km of the mine waste facility?	YES		
Q12	GROUNDWATER Is there a high permeability layer beneath the mine waste facility?			UNKNOWN
Q13	AIR Is the material within the mine waste facility exposed to the wind?	YES		
Q14	DIRECT CONTACT Is the mine waste facility uncovered?	YES		

Table 2. Questions of the third section of the EU MWD Pre-selection Protocol (Q11 to Q14)

2.2.4   *Section 4 of the EU Mine Waste Directive – Pre- selection Protocol*

The section 4 tends to explore 4 major sensitive human and ecosystem receptors The median of the calculated 2 distances is calculated for all threshold limited parameters allowing a threshold estimation representing a 50 % probability of the site falling within the risk limiting distance (median-based threshold) (Abdaal et al. 2013). The Protocol thresholds are based on the Irish regulation for the operation of ponds with respect to quarries (Safe Quarry 2008).

In the final section of the Protocol, the slope of the quarry site was measured using the slope function from the ArcMap 10.5.1 software, the measuring of the distances between water body, agriculture and facilities was performed using the Landsat images. Data about the settlements were obtained via the freely available internet sources.

The results were compared with the European thresholds values taking into the account geographical location, terrain type and physical factors that provided clearer insight into the influence of the mine waste to the environment. Thus, answers on the following questions (Q10 - Q18) indicated that the heap foundation slope is not within the allowed values with the median value of 25˚, water body is distanced about the 900m which regarding the EU legislation marked as an area in danger of contamination from the quarry material. The nearby settlements within the very short distances with the number of 187 inhabitants are a good indicator of the potential contamination risks of the larger community, however the tailing lagoons are not within the area of Natura 2000 but within the very short distance to the agricultural and arable land and permanent crops.
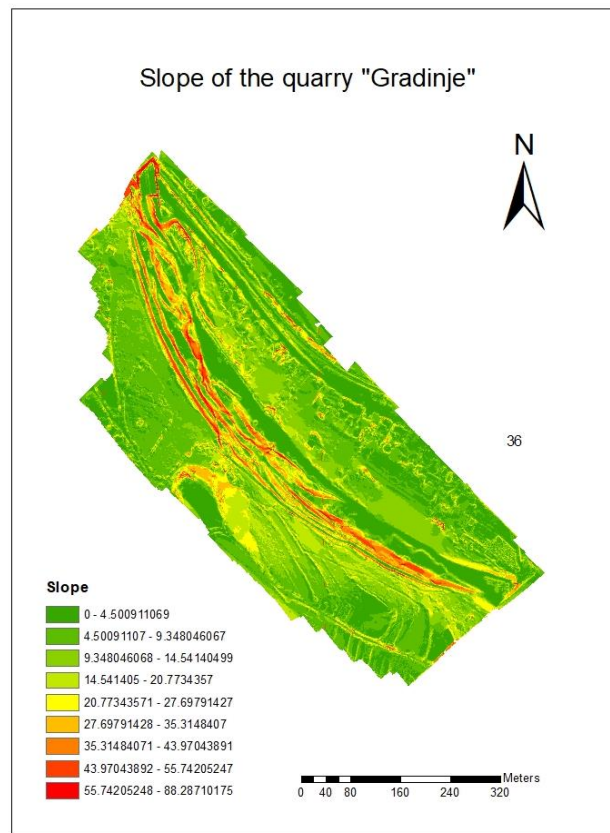
Figure 4. Slope of the quarry "Gradinje"

Question	Issue	EU thresholds (Irish quarry regulations)	Local thresholds (median based)	Basis
Q10	Is the heap foundation greater than a specified slope?	1:12 (5˚)	21˚ -28˚	
Q11	Is there a water body within a specified distance of the facility?	1km	941m 500m	
Q15	Is there a settlement with more than a specified number of people within a specified distance of the facility?	100	187	
Q16	Is the facility within a specified distance from a water body which is at least a specified quality status?	1km	757m	
Q17	Is the facility within a specified distance of a Natura 2000 site?	1km	NO	
Q18	Is the facility within a specified distance of agricultural land?	1km	YES	

Table 3. Questions with threshold values (Q10, Q11 and Q15–18) of the EU MWD Pre-selection Protocol, and the rationale behind each local threshold value

3. DATA AND METHODOLOGY

3.1 Data acquisition

For the purpose of this master thesis, two Landsat images together with the images of Unmanned Aerial Vehicles were used. In order to perform the monitoring of the quarry reveal two Landsat images were used one before and one after the quarry reveal to track the changes. For this reason, two images from the same season were selected, as following the Landsat 7 image (2011, August) and Landsat 8 image (2015, August). Images were chosen from the same season in order to reduce the phenological effect of the vegetation. Those images were obtained from Landsat Enhanced Thematic Mapper Plus (ETM+) which provides images from 2003 and after (based on Landsat 7 satellite), and from Landsat 8 which has been launched on February 11, 2013 for the year of 2015. With the launch of the new Landsat-8 number of composed bands had increased from 8 (Landsat - 7) to 9 bands in Landsat-8. Landsat-8 carries an improved Operational Land Imager (OLI) sensor and the Thermal InfraRed Sensor (TIRS). The OLI sensor provides nine spectral bands with a spatial resolution of 30 m, except the panchromatic band, with spectral resolution of 15 m. The approximate scene size is 170 km north-south by 183 km east-west. For the purpose of this project, 6 bands were kept from both sensors (Blue; Green; Red; NIR; SWIR1 and SWIR2) except from the thermal infrared and panchromatic band which are not essential for my purpose. The spatial resolution amounts to 30 m. The minimum mapping unit of the project is the pixel. This is the unit to which the classification algorithm will be applied.

3.2 Preprocessing of the data

The preprocessing stage of acquired images was conducted in QGIS 2.16 (QGIS Development Team, 2016). The six bands of Landsat - 7 and Landsat - 8 imagery were placed in a single separate data sets for specific years 2011 (before quarry) and 2015 (after quarry has been discovered). The SAVI has been calculated from these individual measurements as follows:

$$SAVI = \frac{(1 + L)(NIR - R)}{(NIR + R + L)}$$

where L is adjustment factor, value of 0.5 was used to minimize the soil brightness; R-stands for the spectral reflectance measurements acquired in the visible (red) region and NIR - stands for the spectral reflectance measurements acquired in the near-infrared region. (https://en.wikipedia.org/wiki/Soil-adjusted_vegetation_index, January 2018) The SAVI difference was performed in order to track the changes in the soil over the period of 5 years. Both images were projected to the WGS_1984_UTM_Zone_34N. One of the preprocessing tasks included the correction of geometric distortions of the Landsat images, for that reason lines were removed using the available options of the QGIS software based on the validity masks for each band.

Secondary, pre-processing of the UAV images was performed using the Agisoft PhotoScan software, which belongs to the category of stand-alone photogrammetric software that allows us to create georeferenced high resolution orthophotos, dense point clouds, and very detailed DSMs/DTMs from the still images. The Agisoft PhotoScan application offers a wide range of output formats such as GeoTIFF, XYZ, Google KML, OBJ, VRML, COLLADA and PDF. These are very desirable formats as they allowed for interoperability with other GIS applications. The obtained UAV images were used as inputs to create the DSM, orthophoto mosaic and 3D point mesh. These products were georeferenced using the GPS data provided. Ground Control Points were as well imported in order to improve the accuracy of the final model. The images were firstly pre-processed by stitching them together and then projecting them using the WGS 84, zone 7 Balkans coordinate system.

3.3 Training assessment

The essential training data for the 6 land cover classes was taken by visual inspection of the original Landsat images, since the resolution of the images was not very high it made this process a demanding task. In total, 200 training samples for six different classes were

used. In process of creating the training set, the most appropriate and representative polygons (field samples) were selected, based on the spectral reflectance and the geometry of certain types of land covers. Specifically and in general terms, I expected the water to be presented as black features; the forest to be depicted with deep red color; the agriculture fields can be discriminated easily by their geometry and the their general proximity to residential areas; rivers, beyond their dark color, they have a unique and very distinct geometry (curved lines crossing large parts of the study area) which make them quite recognizable; bare soil which mostly consist of the quarry area has been shown in high values of DN, since the degree of reflectance is too high; urban areas are shown with greyish (light) color – another characteristic;
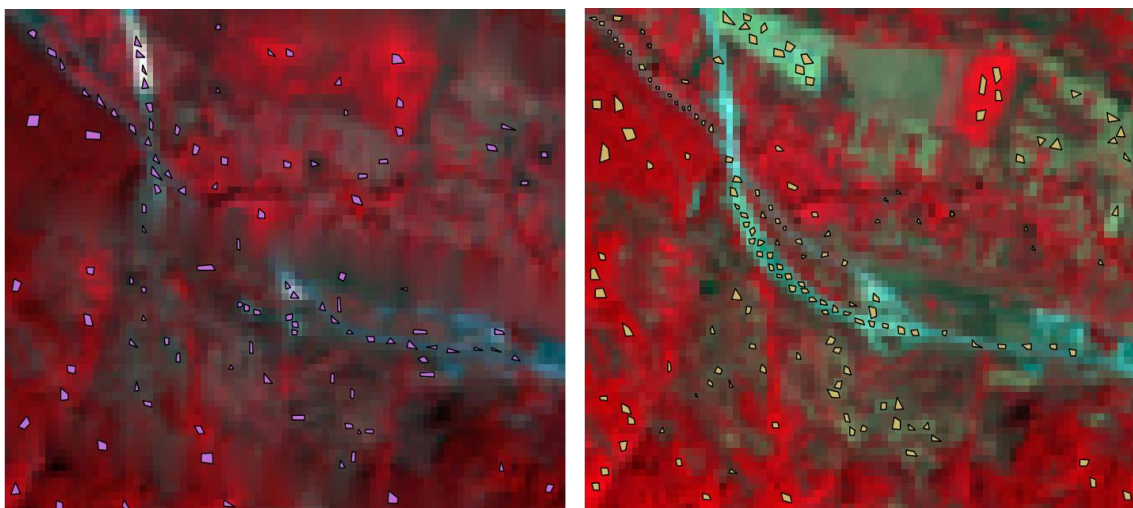
Figure 5. Training samples of the both Landsat images, Author: Nisic, 2018

For assessing the accuracy of the classified map which derived from remote sensing imagery, two techniques are mostly used, namely, overall accuracy and Kappa statistics (Congalton, 1991). Overall accuracy is descriptive statistics, interpreted as proportion of correctly classified pixels in relation to user and producer accuracy (Duro et al., 2012). Kappa statistics is a discrete multivariate technique of accuracy assessment, which functions as a measure of agreement between reference data set and classification results. This is commonly used in assessing statistical difference for classification purposes as well (Congalton, 1991; Duro et al., 2012). Statistical comparison of the classifiers using

only kappa coefficient is not significant (Duro et al., 2012; Foody, 2004). In that case, McNemar's test for paired sample nominal scale data is used by many researchers to assess if there is a statistically significant difference between two classifiers or not (Duro et al., 2012), and is therefore adopted here, so that I could proceed to the appropriate comparisons.

It is crucial to compare the classification algorithm, not only by assessing the accuracy of results, but also by the performance of each classifier during the classification. The use of Receiver Operating Characteristics (ROC) curve has been used in evaluating and comparing algorithms calculating goodness of fit, due to the fact that the simple classification accuracy itself is often a poor metric for measuring performance (Recognition & Bradley, 1997). As ROC curve is a two-dimensional representation of a classifier performance, it is necessary to convert ROC performance to singular scalar value representing expected performance. A common method is to calculate the area under the ROC curve, abbreviated as AUC (Recognition & Bradley, 1997). Evaluation of each classifier's goodness of fit was done by interpreting AUC value and classifiers with highest AUC was considered as high performer.

A brief display of the methodology used in this section is depicted on the following flowchart (Figure 6).

For each classification, a confusion matrix is presented, along with overall accuracy, Kappa statistics, user's accuracy and producer's accuracy. The McNemar's test was performed to assess whether or not there exists statistically significant difference between different classification algorithms. In terms of performance analysis, each models' goodness of fit was evaluated producing area under ROC curve of each algorithm. The ground truth data was used for performing accuracy assessment, statistical comparison, and performance evaluation. In addition, classification time was also estimated as part of comparison of performance efficacy for each classifier.
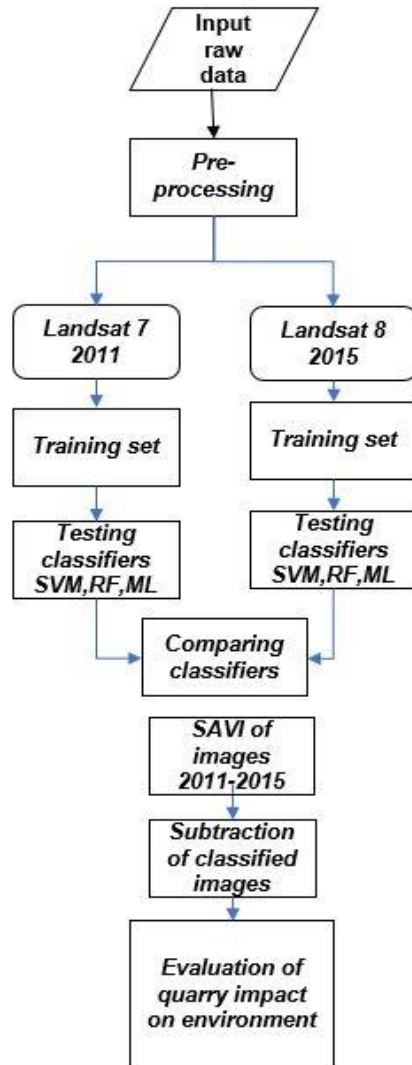
Figure 6. Main methodology of the project

4. EVALUATION OF THE BEST MAPPING APPROACH OF THE TEMPORAL MONITORING OPEN PIT MINE

4.1 Model Building and tuning of classification algorithm

Model building, tuning and accuracy assessment is the repetitive and simultaneous process, which was performed using version 3.3.2 of the 32-bit version of R (R Core Team, 2016) and ArcGIS 10.5.1 software for the case of Maximum likelihood classification. Several add-on packages were retrieved for the implementation of different classifiers in the study area. The classification model using RF algorithm is based on 'randomForest' package (Liaw and Wiener, 2002). For the classification through SVM algorithm, the 'e1071' package was retrieved (Meyer et al., 2017). Maximum likelihood classification was performed in the ArcGIS 10.5.1 software due to some implementation problems in R software. In order to implement classification based on maximum likelihood technique in R, the 'RStoolbox' package was exploited (Leutner and Horning, 2017). Since the suggested package was not giving the results after the long time of execution (approximately 5h), the classification results were obtained using the ArcGIS 10.5.1 software. All classification models performed in R, were developed using the 'caret' packages (Kuhn, 2016) for single consistent environment for training and parameters tuning.

Using the tuning parameters for each algorithm, the classification of the image of the study area was conducted, so that the final land cover maps are to be developed. The 'caret' package was used in order to perform tuning of the parameters for all classification algorithms. The 10-fold cross-validation and 3 repeats were used in order to select the most suitable parameters with highest accuracy of each classifier. Tuning parameters were considered optimal when the algorithms achieve the highest classification accuracy during cross validation. The results were used for accuracy assessment and to compare the performance of the adopted algorithms for land cover classification.

4.2 Maximum Likelihood (ML) based Model

ML is a parametric supervised classification technique which relies on the Bayes theorem. According to Erdas (1999) the algorithm for computing the weighted distance or likelihood D of unknown measurement vector X belong to one of the known classes M_c is based on the Bayesian equation.

$$D = \ln(a_c) - [0.5\ln[|cov_c|)] - [0.5(X - M_c)T(cov_c - 1)(X - M_c)]$$

The unknown measurement vector is assigned to the class in which it has the highest probability of belonging (Otukei & Blaschke 2010). The maximum likelihood classification assumes that spectral values of training pixels are statistically distributed according to a 'multivariate normal probability density function'. For each set of spectral input values, the distance is calculated towards each of the classes. If this distance is smaller than the user-defined threshold value, the class name with the shortest distance is assigned; otherwise, the undefined value is assigned (Toren 2001).
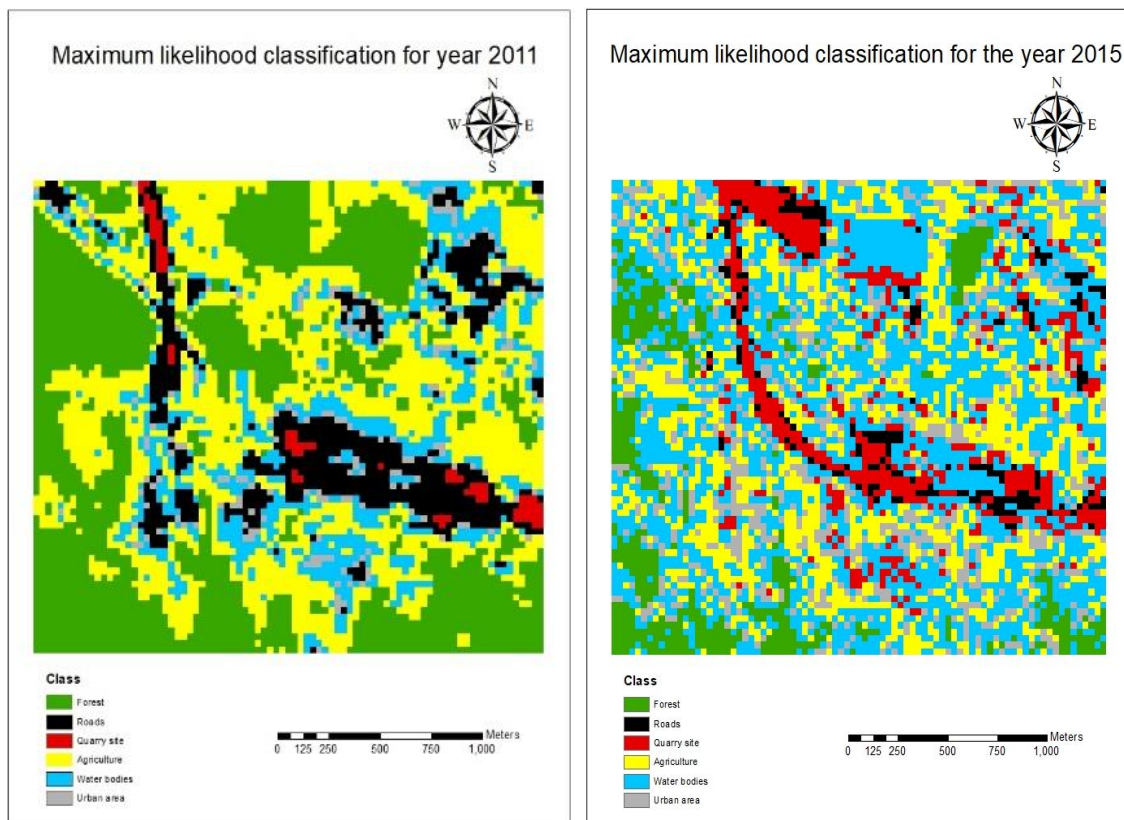
Figure 7. Maximum likelihood classification results for the both years of the examination

In order to establish the separability boundaries between classes, the Jeffries-Matusita (JM) distance can be adopted. This process is more automated without the need of intervening in the structural parameters of the classifier.

The scenario for maximum likelihood algorithm for both years gave quite fuzzy results, as it classified classes of agriculture and forest as a water bodies. It seems that classification of the quarry material performed in a good way, making it easy to distinguish the affected area over the stated period of time.

4.3 Support Vector Machine (SVM)

The theory of the SVM was originally proposed by Vapnik and Chervonenkis (1971) and later discussed in detail by Vapnik (1999). The success of the SVM depends on how well the process is trained. The easiest way to train the SVM is by using linearly separable classes. According to Osuna et al. (1997) if the training data with k number of samples is represented as {Xi, yi}, i = 1, ..., k where $X \in R^N$ is an N-dimensional space and y $\in$ {-1, +1} is a class label then these classes are considered linearly separable if there exists a vector W perpendicular to the linear hyper-plane (which determines the direction of the discriminating plane) and a scalar b showing the offset of the discriminating hyper-plane from the origin (Otukei & Blaschke 2010)**.**

SVM uses kernels functions for establishing the decision rules "to map non-linear decision boundaries in the original data space into linear ones in a high dimensional space". The effectiveness of SVM is primarily relied on the kernel types and parameters determining the limits of decision rules. In this thesis, there was performed SVM classification of the study area using the three types of kernel, as following: linear, radial basis and polynomial. According to the literature review and given results the most accurate results of the classification were achieved using the linear type of the kernel.
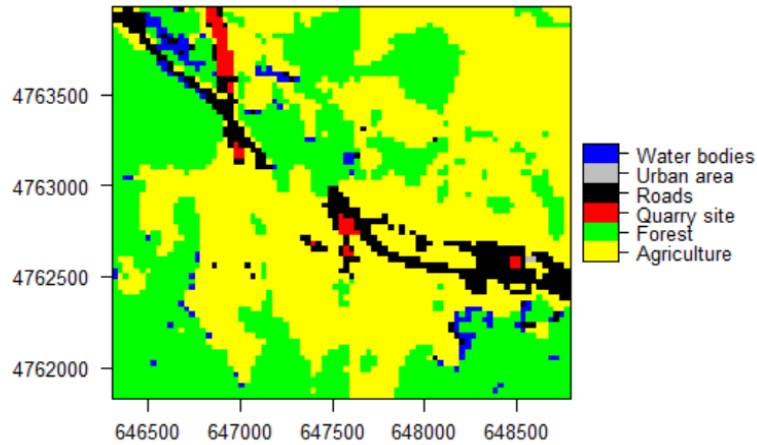
Figure 8. SVM classification result performed on the Landsat 7 image using the linear type of kernel

Based on the visual interpretation of the achieved results the linear type of kernel provided the most accurate and realistic results. The results prove to give a good overview of classification of the all classes, however there is observed the sparse of water pixels on some places, as well as mix of two classes: the quarry sites and roads, but those situations were not having the significant influence on the overall accuracy.
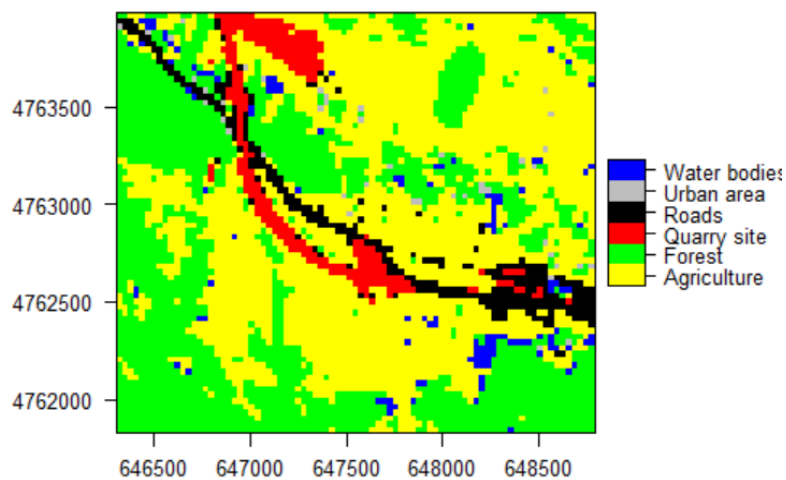
Figure 9. SVM classification result performed on the Landsat 8 image using the linear type of kernel

4.4 Random Forest (RF)

Random forest classifier is an ensemble classifier which consists of a combination of tree classifiers in order to make specific prediction (Rodriguez-Galiano et al., 2012). A RF consists of a combination of classifiers where each classifier contributes with a single vote for the assignation of the most frequent class to the input vector (**x**), $\hat{C}_{rf}^{B} = majorityvote\{\hat{C}_b(\mathbf{x})\}_1^B$, where $\hat{C}_b(\mathbf{x})$ is the class prediction of the bth random forest tree.

The development of a decision tree in RF algorithm is mainly depended upon the choice of attribute selection at each split in the tree building process (mtry) and the number of trees (ntree) (Belgiu & Drăgu, 2016; Pal, 2005). Default number of tree (500) in R packages was selected because it is reported that the errors are stabilized before this number of classification trees and have negligible influence on overall classification accuracy (Belgiu & Drăgu, 2016; Duro et al., 2012). Mtry is believed to have "sensitive" influence on the performance of RF classifier (Duro et al., 2012). Usually, the value of Mtry is set to be the square root of the number of predictor variables used (Belgiu & Drăgu, 2016).

According to the Breiman's research and suggestion of the random forest as a innovative and efficient approach, here are given some of the advantages for it application in remote sensing:

- It runs efficiently on large data bases.
- It can handle thousands of input variables without variable deletion.
- It gives estimates of what variables are important in the classification.
- It generates an internal unbiased estimate of the generalization error (oob error).
- It computes proximities between pairs of cases that can be used in locating outliers.
- It is relatively robust to outliers and noise.

- It is computationally lighter than other tree ensemble methods (e.g. Boosting)(Rodriguez-Galiano et al. 2012).
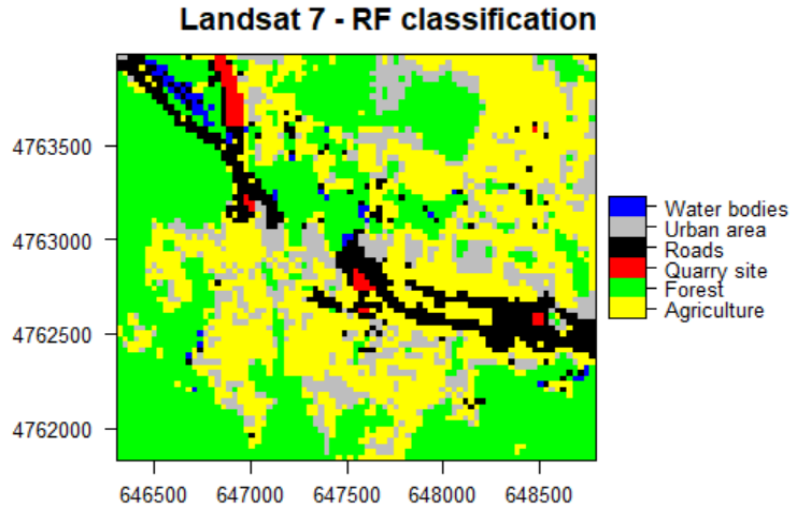
Figure 10. RF classification result performed on the Landsat 7 image

In general terms, the results of the random forest classifier seem to be representative with the land cover of the study area. Definitely, this argument will be examined in the accuracy assessment section. The quarry areas cover a significant part of both dominant land covers, namely, the agricultural and forested lands as present in the reality. But, the map of changes will delineate the quarry areas in more detail. In the classification result for the year 2011 (Figure 8.), there are observed very obvious errors that should be mentioned. First of all, there are some urban areas on the extreme northern, central and southern parts of the study domain which do not agree with the ones in the reality. From the satellite images there could be concluded that the mentioned area consists mainly from the agriculture and forest land cover types.

Moreover, there were observed the errors regarding the classification of the two classes: forest and agriculture, in the north eastern part of the study area of Landsat 8 image (Figure 9.)
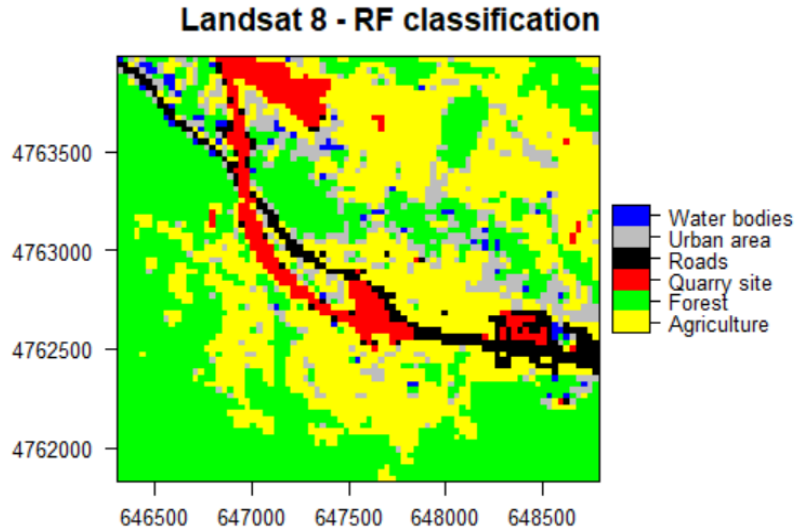
Figure 11. RF classification result performed on the Landsat 8 image

4.5 Comparison of the classifiers

4.5.1 Theoretical framework

Machine learning methods have been applied successfully for classification, and their use is increasing due to the availability of faster computing capabilities. Two machine learning methods which have given good results in previous investigations are Random Forest (RF) and Support Vector Machine (SVM) (Piragnolo et al. 2017). The analysis of the performance of SVM and RF has been done using three accuracy metrics: Kappa index, classification accuracy and classification error. The classification accuracy A_i of an individual program i depends on the number of samples correctly classified (true positives plus true negatives) and is evaluated by the formula:

$$A_i = \frac{t}{n} \cdot 100$$

where t is the number of sample cases correctly classified, and n is the total number of sample cases. The classification error Ei of an individual program i depends on the

number of samples incorrectly classified (false positives, plus false negatives) and is evaluated by the formula:

$$E_i = \frac{f}{n} \cdot 100$$

where f is the number of sample cases incorrectly classified, and n is the total number of sample cases (https://www.gepsoft.com/gepsoft/APS3KB/Chapter09/Section2/SS02.htm).

All of them were created using the different ways: with K-fold cross validation, using the validation test set. Validation with K-fold and with the validation dataset show SVM give better results, but RF prove to be more performing when training size is larger. Classification error and classification accuracy follow the trend of Kappa index (Piragnolo et al. 2017).

SVM is intrinsically a contemporary binary machine learning algorithm that has been applied extensively in several disciplines. However, applications of binary classification are very limited especially in RS land cover classification where most of the classification problems involve more than two classes (Melgani and Bruzzone, 2004). According to the literature review, machine learning algorithms gave more satisfying results over the classical approaches.

4.5.2 Sensitivity analysis

ML classification model is automated without structural parameter, so there was no need for tuning of the parameters. The overall classification accuracy achieved by ML amounts to 61.2% for the year 2011 and 42.8% for 2015. On the other hand, the highest classification accuracy value (86.0%) for the year before the discovery of the quarry and 73.5% was obtained with RF classifier for the image of the year, 2015. SVM classifier gave a slightly lower classification accuracy results (82.8%) in case of the Landsat 7 image, but for the Landsat 8 image it shows to perform better than RF model obtaining the overall classification accuracy of 80%. Regarding the SVM model, the combination of three critical parameters (cost, gamma and type of kernel) were adjusted accordingly.

The optimal parameters for cost and gamma were (cost = 1, gamma = 1), in case of kernel type the highest classification accuracy for both years was obtained using the linear type of kernel. Models with optimal parameters were used to produce the classified map, associated accuracy assessment and map comparisons. The two machine learning algorithms produced a visually more coherent map than ML algorithm. The major difference among the three classification algorithms was observed on the southern and central part of the study domain. Particularly, RF algorithm classified urban area in the central and southern part of the study area; whereas SVM algorithm classified it as agriculture.

The scenario for maximum likelihood algorithm is different, central part of the study area as well some parts in the north and south were classified as water and roads area which makes it drastically different compared with the SVM and RF classifier. In addition, RF and ML algorithm classified the northern part of the area as a urban area whereas SVM classified it as agriculture. It is also obvious that ML classifier suffered in a significant degree compared with the other two algorithms.

4.5.3 Accuracy Assessment and statistical comparison

Accuracy assessment consisted one of the last stages for each classification procedure, was performed to evaluate the efficiency of each model prediction against the ground truth data collected using the Google Earth software. Table 4 shows the detailed confusion matrix for the most accurate classifier (RF) based on the training data with overall accuracy, user accuracy, producer accuracy and kappa statistics.

Random Forest classifier (2011)									
Class	**1**	**2**	**3**	**4**	**5**	**6**	Total	U_Accuracy	Kappa
1	92.000	5.000	1.000	0.000	1.000	1.000	100.000	0.920	0.000
2	7.000	80.000	0.000	0.000	0.000	0.000	87.000	0.920	0.000
3	0.000	2.000	6.000	2.000	0.000	0.000	10.000	0.600	0.000
4	5.000	1.000	1.000	6.000	0.000	0.000	13.000	0.462	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	3.000	1.000	0.000	0.000	0.000	1.000	5.000	0.200	0.000
Total	107.000	89.000	8.000	8.000	1.000	2.000	215.000	0.000	0.000

	1	2	3	4	5	6	Total	U_Accuracy	Kappa
P_Accuracy	0.860	0.899	0.750	0.750	0.000	0.500	0.000	0.860	0.000
Kappa	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.766

Table 4. Confusion matrix of the RF classification for the year of 2011

Random Forest classifier (2015)									
Class	**1**	**2**	**3**	**4**	**5**	**6**	Total	U_Accuracy	Kappa
1	67.000	2.000	1.000	2.000	0.000	0.000	72.000	0.931	0.000
2	19.000	69.000	0.000	0.000	0.000	1.000	89.000	0.775	0.000
3	0.000	0.000	6.000	2.000	1.000	0.000	9.000	0.667	0.000
4	3.000	0.000	0.000	5.000	1.000	0.000	9.000	0.556	0.000
5	13.000	6.000	0.000	0.000	0.000	0.000	19.000	0.000	0.000
6	1.000	0.000	0.000	1.000	0.000	0.000	2.000	0.000	0.000
Total	103.000	77.000	7.000	10.000	2.000	1.000	200.000	0.000	0.000
P_Accuracy	0.650	0.896	0.857	0.500	0.000	0.000	0.000	0.735	0.000
Kappa	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.585

Table 5. Confusion matrix of the RF classification for the year of 2015

The general trend of overall classification accuracy shows that ML algorithm presented the lowest overall classification accuracy (61.4% and 41.0%) respectively for both years 2011 and 2015, followed by SVM (82.8% and 80.0%) and RF algorithm (86.0% and 73.5%). Similar trend was observed for kappa statistics, where ML, SVM and RF algorithm yielded respectively to 42.8% and 23.8%, 71.2% and 66.7% and 76.6% and 58.5% respectively the years of 2011 and 2015. At the same time, the respective percentages for user's and producer's accuracy are high (above 70%) from all classification algorithms regarding the agriculture, forest, quarry sites and roads (APPENDIX). Discrimination of the urban area and agriculture fields was difficult, since the area covered with the urban was relatively small and it influenced stratified random sampling of the points that were later used for accuracy assessment.

Finally, concerning the comparison between different classification predictions, the McNemar test indicated that the observed difference between the two machine learning algorithms (RF and SVM) was not statistically significant ($p = 0.2278 > 0.05$). However, it should be emphasized that the further examination of the comparison between the machine learning algorithms and classical approach of ML was not performed due to drastically different classification accuracy.

4.5.4 Performance evaluation

A ROC curve along with AUC value for each algorithm was generated to evaluate the overall performance in term of goodness of fit of the classifiers, as showed in Figure 2.
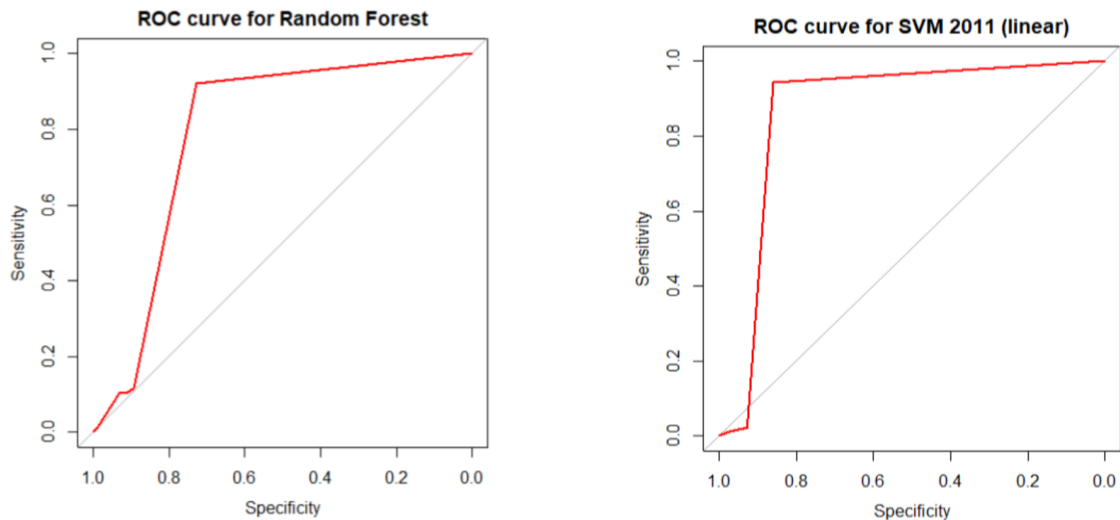
Figure 12. ROC curve for support vector machine and random forest classifier

Based on the exploration of the ROC curve, it may be noticed that RF algorithm provided very satisfying results in terms of performance, which is presented with the shape of the ROC curve. The goodness of fit of SVM classifier is comparable to RF, presenting almost similar ROC curve with RF's. Furthermore, the general trend of AUC value confirms the value under the curve for SVM algorithm (AUC = 0.8376) and RF algorithm (AUC = 0.8716).

As a last factor/final way for comparison of the classifiers, time needed for algorithm to complete the classification process was taken into the account. The results showed a very slight difference between two of machine learning algorithms as follows: RF (7seconds, CPU time) and SVM (5 seconds, CPU time).

4.5.5 The impact of the quarry on the environment during the period of 5 years

The SAVI index was calculated for each year separately and afterwards the subtraction of the unaffected area from the year after the incident (2015) was performed. Creation of the SAVI ancillary data has the purpose of distinguishing the areas affected with the quarry sites. With the first identification of the images, there could be observed that agricultural and forested area before the quarry reveal has a greater distribution in the central and northern part of the study area. On the second image, there is very noticeable presence of the two sites where open pit mine excavation took place.
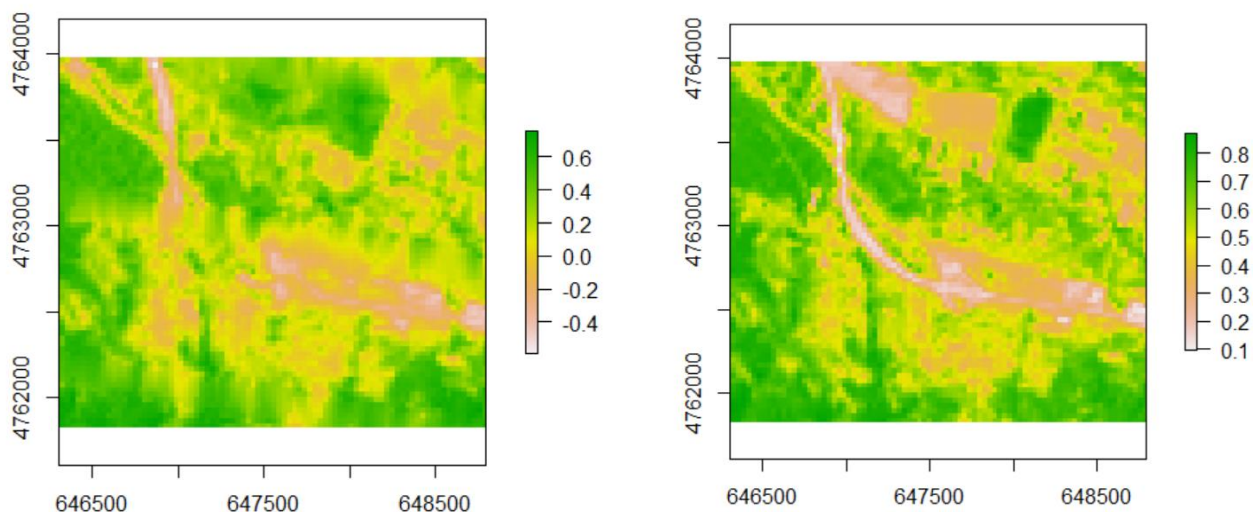
Figure 13. SAVI index for the both situations, before and after the quarry has been revealed

After the subtraction of the images, map of changes was created from which it can be observed that the quarry event has affected a significant amount of the total surface of the study area, making a damage to the extensive areas of forested and agricultural land, as well as on the urban area. The massive implications of this event are related with: i) the human health complication to the residents from the nearby settlements; ii) the huge environmental damage destructing a valuable natural resource like forest and the respecting flora and fauna; iii) the contamination of many cultivation fields; iv) extensive soil degradation and the resulting escalating of river contamination risk;
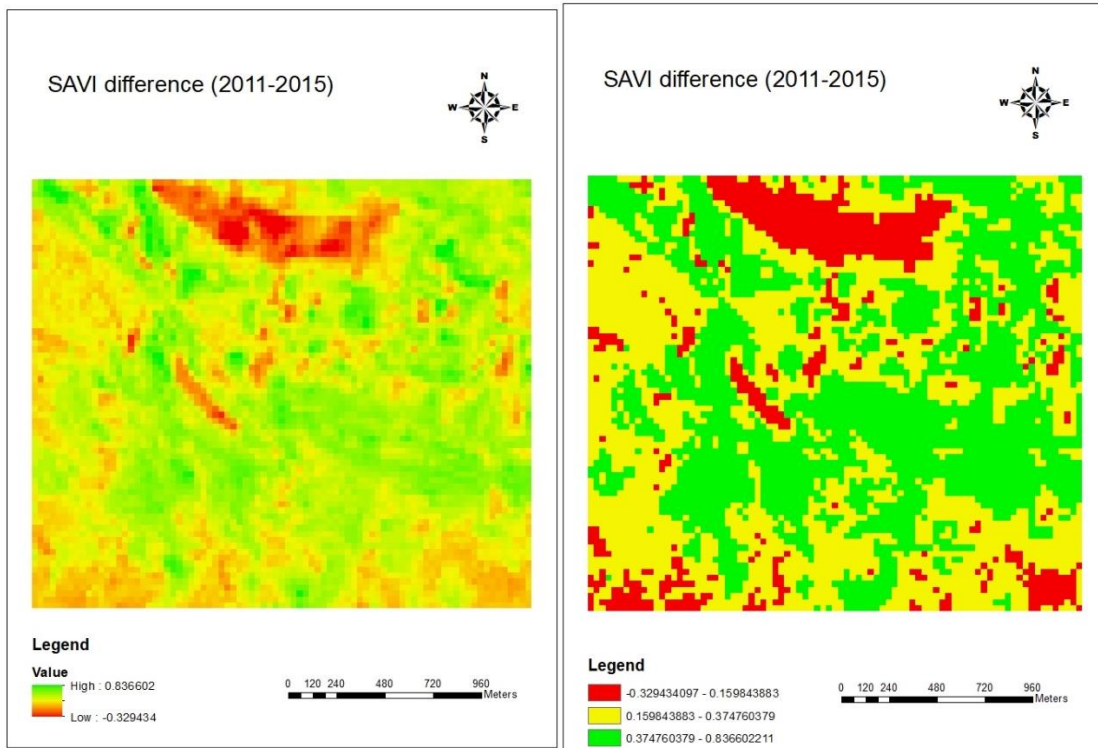
Figure 14. Map of changes in the land cover (2011-2015)

The last part of the change detection constitutes the quantification of the area affected with quarry reveal. For the purpose of detecting eventual changes in the landcover over the time of 5 years, confusion matrices of both studied years were normalized and compared. It was observed that there is decrease of the agriculture and cultivation fields by 10.51%, as well as slightly increase of the forest class within the study area by 4.05% which may be related with the classification errors. The results seem meaningful and in compliance with the SAVI difference map.

5. DISCUSSION

Generally, the classification of the land cover maps using the RF and SVM algorithms provided comparable and more coherent representation of the land cover maps within the study area, while on the other hand ML classification method produced significantly different output compared with the other two classifiers, which made its result less discernible.

In terms of the classification accuracy taking into the account overall accuracy and kappa statistics between the algorithms, there could be noticed a slight improvement in accuracy (1.65%-average value) when comparing the machine learning algorithms (SVM and RF). However, a significant difference in improvement of the accuracy was observed when choosing the machine learning algorithms over the classical parametric algorithm (ML) as follows by 30.2% for SVM and by 28.55% for RF classifier. McNemar's test revealed that there is no statistical significant difference between classification accuracies achieved by applied machine learning algorithms (RF and SVM). Nonetheless, classification accuracy achieved with RF and SVM according to the MCNemar's test was showing statistically significant difference ($p < 0.05$) when compared with ML classifier.

Other studies have indicated that both RF and SVM can achieve similar classification accuracies outperforming the corresponding accuracy of ML. Duro et al. (2012) reported that pixel based classification using RF and SVM algorithms produced statistically similar overall classification accuracy classifying agricultural landscapes. Otukei & Blaschke (2010) found that SVM performed better than classical ML algorithm. In the same context, Sesnie et al. (2010) performed classification of rainforest types using SVM and RF classifiers and found to perform comparatively well.

When compared using the ROC curve (AUC) value, it is recorded/noticed that there is small change in goodness of fit of the models using the RF model comparatively to SVM; a greater difference of goodness in fit was produced by both RF and SVM compared to ML. However, RF and SVM algorithms showed better results in time execution for the whole process of image classification (modeling and predicting) than the ML.

It is argued that classification accuracy is a critical parameter for selecting the appropriate classifier for image classification. But in such circumstances where there is no statistically significant difference in classification accuracy between RF and SVM algorithms, other preferences may take precedence. RF is too simple in parameter tuning as it only should change one parameter for optimizing algorithm; whereas in SVM classifier should be supplied with combination of two important parameters for optimization. Moreover, RF is preferred for its accessibility to variable importance (Sesnie et al., 2010). The performance of RF in term of time execution is also slightly higher than that of SVM's.

6. CONCLUSIONS

The main environmental effects caused with the discovery of the quarry are affecting the agricultural land by modification and pollution of the nearby, pre-existing ecosystems. With regards to the prevailing environmental legislation and its enforcement, there is total lack of efforts in monitoring, rehabilitation, restoration or post-mining programmes for minimization of adverse environmental impacts (Lad R. J. and Samant J. S. 2014). The most obvious engineering impact of quarrying is change in geomorphology and conversion of land use, with the associated change in visual scene (Langer 2001). Most engineering impacts can be controlled, mitigated, kept at tolerable levels and restricted to the immediate vicinity of the aggregate operation by employing responsible operational practices that use available engineering techniques and technology (Langer 2001). Major changes in the land cover as well as increase of the contaminated mine sites in Europe has influenced the development of the EU environmental legislation in order to minimize potential risks of such incidents. EU commission has developed the legislation named EU MWD (Mine Waste Directive) pre-selection Protocol which consist of the 18 simple questions about contamination source, pathways and receptors. The study area has been tested for the contamination and compliance of the quarry sites with the legislation using the simple set of questions, and the results obtained show the unexpected outcome of unknown parameters regarding the engineering facility. Likewise the accumulated number of YES responses has indicated the presence of the contamination sources, pathways and receptors, which later was monitored using the appropriate algorithm for the classification of the land cover and quarry site.

Pixel based classification of land cover using classical ML and two machine learning algorithms i.e. SVM and RF algorithms was performed. Machine learning algorithms provide a lot of potential in pixel-based land cover classification over classical approach, as machine learning approach produced classification accuracies that had statistically significant difference compared to classical approach. But there is no statistical significant difference in classification accuracy between the two adopted machine learning algorithms. The performance of machine learning algorithm-based goodness of fit is also

higher than classical approach. Visual inspection of thematic products resulted from the three algorithms shows that they are capable of producing land cover map with acceptable classification accuracies; however, ML classifier showed quite fuzzy results. So, machine learning algorithms seem a bit superior than classical ML for the classification of the open pit mines within this study area.

BIBLIOGRAPHIC REFERENCES

ABDAAL, A., JORDAN, G. & SZILASSI, P., 2013. Testing contamination risk assessment methods for mine waste sites. *Water, Air, and Soil Pollution*, 224(2).

AHMAD, A. & QUEGAN, S., 2012. Analysis of Maximum Likelihood Classification on Multispectral Data. *Applied Mathematical Sciences*, 6(129), pp.6425–6436.

ANON, 2016. Analysis of Object Discrimination and characterisation from Images for Classification through machine learning (February).

BAJIC, M., IVELJA, T. & BROOK, A., 2017. Developing a Hyperspectral Non-Technical Survey for Minefields via UAV and Helicopter. *Journal of Conventional Weapons Destruction Journal of Conventional Weapons Destruction Issue Journal of Conventional Weapons Destruction*, 21(21.1), pp.51–63. Available at: http://commons.lib.jmu.edu/cisr-journal/vol21/iss1/11.

BENJAMIN, A. & LEUTNER, M.B., 2017. Package " RStoolbox ."

BENTO, A.P. et al., 2013. Classification and Regression by randomForest. *Nucleic Acids Research*, 5(1), pp.983–999. Available at: http://www.cs.colorado.edu/department/publications/reports/docs/CU-CS-954-03.pdf%5Cnpapers2://publication/uuid/F28DC17B-D961-4E66-9DEF-B940467A5068%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3223189&tool=pmcentrez&rendertype=abstract%5Cn.

BEREGOVOI, D. V., YOUNES, J.A. & MUSTAFIN, M.G., 2017. Monitoring of Quarry Slope Deformations with the Use of Satellite Positioning Technology and Unmanned Aerial Vehicles. *Procedia Engineering*, 189(May), pp.737–743. Available at: http://dx.doi.org/10.1016/j.proeng.2017.05.116.

BRADLEY, A., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), pp.1145–1159.

BRERETON, R.G. & LLOYD, G.R., 2010. Support vector machines for classification and regression. *The Analyst*, 135(2), pp.230–267.

CAETANO, M., 2013. Theory on Land Use / Cover & Change Detection Theory on Land Cover and Land Use Monitoring D4T1b , pp.1–87.

CANDADE, N., 2004. Multispectral Classification of Landsat Images: a Comparison of Support Vector Machine and Neural Network Classifiers. *ASPRS Annual Conference Proceeding*, (May), p.12.

CHANDARAN, S., 2015. Impact Of Granite Quarry on Human Life and Environment A Case Study of Vellarada Panchayat of Thiruvanantahpuram District, Kerala. *on*

Climate Change and Developing World, (March). Available at: https://works.bepress.com/sarath_chandran/2/download/.

CHARBONNIER, P., 2001. Management of mining, quarrying, and ore-processing waste in the European Union. *Brgm*, (December), pp.1–88. Available at: http://ec.europa.eu/environment/waste/studies/mining/0204finalreportbrgm.pdf.

CONGALTON, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), pp.35–46.

D. GONZALES-AGUILERA et al., 2012. 3D Modelling and Accuracy Assessment of Granite Quarry Using Unmmanned Aerial Vehicle. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, I-3(September), pp.37–42. Available at: http://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/I-3/37/2012/.

Dr. POONAM S., K.R., 2016. Analysis of Object Discrimination and characterisation from Images for Classification through machine learning. *International Journal of Computer Architecture and Mobility (ISSN 2319-9229)*, 4 Issue(January 2016), p.7.

DUDLEY, N., 2008. *Guidelines for Protected Area Management Categories*, Available at:http://books.google.com/books?hl=en&lr=&id=4gi3vtvRJ9EC&oi=fnd&pg=PT1 1&dq=Guidelines+for+Protected+Area+Management+Categories&ots=WeYG-lBWPG&sig=d1K4MqXZ2ga7HoAzAUomkeao-JQ.

DURO, D.C., FRANKLIN, S.E. & DUBE, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, 118, pp.259–272. Available at: http://dx.doi.org/10.1016/j.rse.2011.11.020.

EL MEOUCHE, R. et al., 2016. Uav Photogrammetry Implementation To Enhance Land Surveying, Comparisons and Possibilities. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W2(October), pp.107–114. Available at: http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W2/107/2016/.

FILIPOVA, S., FILIPOV, D. & RAEVA, P., 2016. Creating 3D Model of an Open Pit Quarry By Uav Imaging and Analysis in Gis. *Cartography-Gis.Com*, (June), pp.13–17. Available at: http://cartography-gis.com/docsbca/iccgis2016/ICCGIS2016-67.pdf.

FOODY, G.M., 2004. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering & Remote Sensing* , 70(5), pp.627–633.

FRIGUI, H., GADER, P.D. & CHAMSEDDINE BEN ABDALLAH, A., 2008. A generic framework for context-dependent fusion with application to landmine detection. *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIII*, 6953, 6953(December), p.10. Available at: http://chams.mouhendis.com/research/papers/Fusion_FriguiGader.pdf%5Cnhttp://p roceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1332186.

GARCIA, M. et al., 2011. Multispectral and LiDAR data fusion for fuel type mapping using Support Vector Machine and decision rules. *Remote Sensing of Environment*, 115(6), pp.1369–1379. Available at: http://dx.doi.org/10.1016/j.rse.2011.01.017.

GHIMIRE, B., ROGAN, J. & MILLER, J., 2010. Contextual land-cover classification: Incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters*, 1(1), pp.45–54.

HE, J. et al., 2015. A comparison of classification algorithms using Landsat-7 and Landsat-8 data for mapping lithology in Canada???s Arctic. *International Journal of Remote Sensing*, 36(8), pp.2252–2276.

HUANG, C., DAVIS, L.S. & TOWNSHEN, J.R.G., 2002. An assessment of support vector machines for land cover classiŽ cation. *International Journal of Remote Sensing*, 23(4), pp.725–749.

IMMITZER, M., ATZBERGER, C. & KOUKAL, T., 2012. Tree species classification with Random forest using very high spatial resolution 8-band worldView-2 satellite data. *Remote Sensing*, 4(9), pp.2661–2693.

JORDAN, G., 2014. Environmental modelling and spatial landscape analysis for the contamination, pp.1–104.

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

KORUYAN, K. et al., 2012. Remote sensing in management of mining land and proximate habitat. *Journal of the Southern African Institute of Mining and Metallurgy*, 112(7), pp.667–672.

LAD R. J. and SAMANT J. S., 2014. Environmental and social imapcts oof stone quarrying-a case study of Kolhapur district, International Journal of Current research, p.pp.5664-5669.

LANGER, W.H., 2001. Potential environmental impacts of quarrying stone in karst–A literature review. *US Geological Survey*, p.34. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Potential+Enviro nmental+Impacts+of+Quarrying+Stone+in+Karst-#0.

MAS, J.-F., 1999. Monitoring land-cover changes: a comparison of change detection techniques. *International Journal of Remote Sensing*, 20(1), pp.139–152.

MELGANI, F. & BRUZZONE, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. and Remote Sens.*, 42(JANUARY 2004), pp.1778–1790. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1323134.

MEYER, D. et al., 2017. *R Package e1071 Version 1.6-8*, Available at: https://cran.r-project.org/web/packages/e1071/e1071.pdf.

MOLNAR, A., 2016. Volume analysis of surface formations on the basis of aerial photographs taken by drones Faculty of Economy 2 3D model creation based on the. *International Journal of Signal Processing*, 1(ISSN: 2367-8984), pp.152–159.

NITZE, I., SCHULTHESS, U. & ASCHE, H., 2012. Comparison of Machine Learning Algorithms Random Forest , Artificial Neural Network and Support Vector Machine To Maximum Likelihood for Supervised Crop Type Classification. , (February 2016).

OKWUASHI, O., 2014. One against all remote sensing, pp.304–316.

OTUKEI, J.R. & Blaschke, T., 2010. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, 12(SUPPL. 1).

OYINLOYE, M.A. & AJAYI, G.E., 2015. The Journal of Macro Trends in Energy and Sustainability ENVIRONMENTAL IMPACT ASSESSMENT OF QUARRY ACTIVITIES IN OBA-ILE , ONDO STATE , NIGERIA. , 3(1), pp.31–45.

PAPER, R. et al., 2015. Geo-mapping and Visual Stitching to Support Landmine Detection Using a Low- cost UAV Regular Paper. *International Journal of Advanced Robotic Systems*, (May 2015), p.11.

PIRAGNOLO, M., MASIERO, A. & PIROTTI, F., 2017. Comparison of Random Forest and Support Vector Machine classifiers using UAV remote sensing imagery. *Geophysical Research Abstracts EGU General Assembly*, 19(iii), pp.2017–15692. Available at: http://meetingorganizer.copernicus.org/EGU2017/EGU2017-15692-1.pdf.

QIAN, Y. et al., 2015. Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. *Remote Sensing*, 7(1), pp.153–168.

RACZKO E. & ZAGAJEWSKI, B., 2017. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on

airborne hyperspectral APEX images. *European Journal of Remote Sensing*, 50(1), pp.144–154. Available at: https://www.tandfonline.com/doi/full/10.1080/22797254.2017.1299557.

RICHARDS, J. & JIA, X., 2006. *Remote Sensing Digital Image Analysis*, Available at: http://books.google.com/books?id=4PB5vhPBdJ4C.

RODRIGUEZ-GALIANO, V.F. et al., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(1), pp.93–104. Available at: http://dx.doi.org/10.1016/j.isprsjprs.2011.11.002.

SAROGLU, E. et al., 2005. Environmental Impact Analyses of Quarries Located on the 1Asian Side of Istanbul Using Remotely Sensed Data. *XXII International Cartographic Conference (ICC), Corunna, Spania*, (July), pp.11–16.

SIMP, A.X. V & REMOTO, S., 2011. Optimizing land cover change detection using combined pixel-based and object-based image classification in a mountainous area in Mexico. *Anais XV Simpósio Brasileiro de Sensoriamento Remoto*, pp.6556–6563.

TOREN, T., 2001. Assessment of Open Pit Coal Mining Impacts Using Remote Sensing : A Case Study from Turkey. , pp.461–466.

TOWL, G., 2005. Risk assessment. *Evidence-Based Mental Health*, 8(4), pp.91–93. Available at: http://ebmh.bmj.com/cgi/doi/10.1136/ebmh.8.4.91.

VOROVENCII, I., 2011. The Assessment of the Impact on the Environment of the Limestone Quarries Using Satellite Images. *Environmental Engineering & Management Journal (EEMJ)*, 10(10), pp.1511–1522. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=8gh&AN=69668877&site =ehost-live.

WEFERS S., ATORF P., K.J., UAV photogrammetry and 3D analyses of CH sites The millstone quarry district of Mayen (DE) as a case study. *International Conference on Cultural Heritage and New Technologies | Vienna | 2015*, pp.135–141.

APPENDIX

START

1. KNOWN IMPACT
Is the mine waste facility known to have an incident which had a serious impact on human health or environment?

YES

NO

2A. SOURCE - contents
Is the mine waste facility a potential **source** of *pollutants*?

YES

NO

2B. SOURCE - stability
Is the **source** potentially physically *unstable*?

NO

YES

3. PATHWAY
Are there potential **pathways** for possible contamination to reach potential receptors?

NO

YES

4. RECEPTOR
Are there known potential **receptors** in the area?

NO

YES

END POINT 1
NO NEED TO EXAMINE FURTHER

END POINT 2
EXAMINE FURTHER
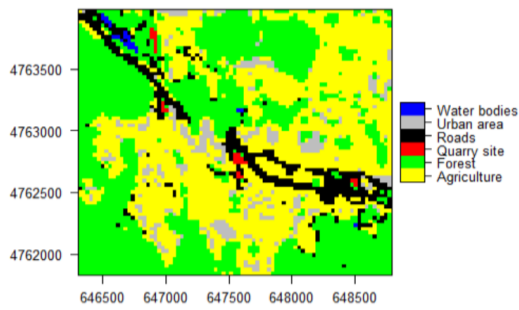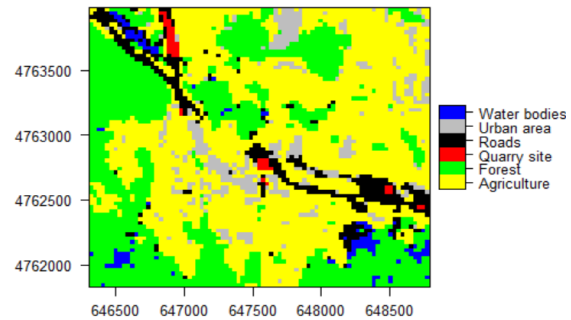
SVM polynomial(2011)									
Class	**1**	**2**	**3**	**4**	**5**	**6**	Total	U_Accuracy	Kappa
1	83.000	12.000	0.000	2.000	1.000	1.000	99.000	0.838	0.000
2	4.000	66.000	0.000	0.000	0.000	0.000	70.000	0.943	0.000
3	0.000	2.000	6.000	2.000	0.000	0.000	10.000	0.600	0.000
4	5.000	3.000	0.000	4.000	0.000	0.000	12.000	0.333	0.000
5	13.000	1.000	2.000	0.000	0.000	0.000	16.000	0.000	0.000
6	2.000	5.000	0.000	0.000	0.000	1.000	8.000	0.125	0.000
Total	107.000	89.000	8.000	8.000	1.000	2.000	215.000	0.000	0.000
P_Accuracy	0.776	0.742	0.750	0.500	0.000	0.500	0.000	0.744	0.000
Kappa	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.595

SVM radial (2011)									
Class	**1**	**2**	**3**	**4**	**5**	**6**	Total	U_Accuracy	Kappa
1	73.000	10.000	1.000	1.000	1.000	1.000	87.000	0.839	0.000
2	16.000	77.000	3.000	2.000	0.000	0.000	98.000	0.786	0.000
3	0.000	0.000	4.000	1.000	0.000	0.000	5.000	0.800	0.000
4	7.000	1.000	0.000	4.000	0.000	0.000	12.000	0.333	0.000
5	11.000	1.000	0.000	0.000	0.000	0.000	12.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	0.000
Total	107.000	89.000	8.000	8.000	1.000	2.000	215.000	0.000	0.000
P_Accuracy	0.682	0.865	0.500	0.500	0.000	0.500	0.000	0.740	0.000
Kappa	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.571

SVM linear(2011)									
Class	**1**	**2**	**3**	**4**	**5**	**6**	Total	U_Accuracy	Kappa
1	87.000	7.000	1.000	0.000	1.000	1.000	97.000	0.897	0.000
2	12.000	78.000	0.000	0.000	0.000	0.000	90.000	0.867	0.000
3	0.000	2.000	6.000	2.000	0.000	0.000	10.000	0.600	0.000
4	5.000	1.000	1.000	6.000	0.000	0.000	13.000	0.462	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	3.000	1.000	0.000	0.000	0.000	1.000	5.000	0.200	0.000
Total	107.000	89.000	8.000	8.000	1.000	2.000	215.000	0.000	0.000
P_Accuracy	0.813	0.876	0.750	0.750	0.000	0.500	0.000	0.828	0.000
Kappa	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.712

Landsat 8 - SVM classification (radial basis)
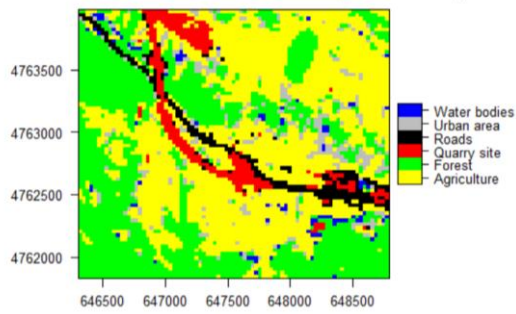
Landsat 8 - SVM classification (polynomial)
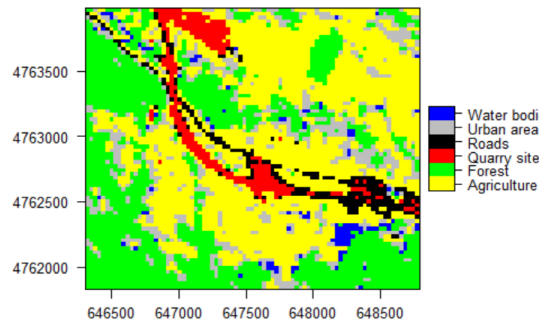
ROC curve for SVM 2011 (polynomial)
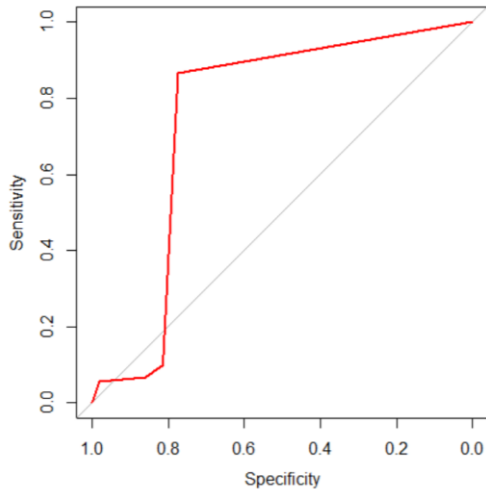
ROC curve for SVM (radial basis)
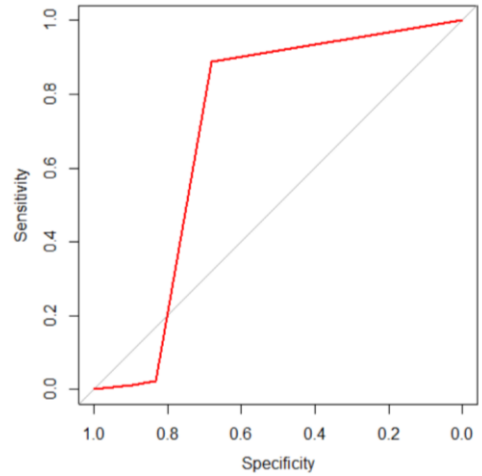
SVM polynomial (2015)									
Class	1	2	3	4	5	6	Total	U_Accuracy	Kappa
1	73.000	9.000	0.000	2.000	0.000	0.000	84.000	0.869	0.000
2	11.000	52.000	0.000	0.000	0.000	1.000	64.000	0.813	0.000
3	0.000	0.000	6.000	2.000	1.000	0.000	9.000	0.667	0.000
4	4.000	0.000	1.000	5.000	1.000	0.000	11.000	0.455	0.000
5	11.000	13.000	0.000	1.000	0.000	0.000	25.000	0.000	0.000
6	2.000	2.000	0.000	0.000	0.000	0.000	4.000	0.000	0.000
Total	101.000	76.000	7.000	10.000	2.000	1.000	197.000	0.000	0.000
P_Accuracy	0.723	0.684	0.857	0.500	0.000	0.000	0.000	0.690	0.000
Kappa	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.524

SVM linear (2015)									
Class	**1**	**2**	**3**	**4**	**5**	**6**	Total	U_Accuracy	Kappa
1	81.000	9.000	0.000	2.000	0.000	0.000	92.000	0.880	0.000
2	16.000	65.000	0.000	0.000	0.000	1.000	82.000	0.793	0.000
3	2.000	0.000	7.000	0.000	1.000	0.000	10.000	0.700	0.000
4	2.000	0.000	0.000	7.000	1.000	0.000	10.000	0.700	0.000
5	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
6	2.000	3.000	0.000	0.000	0.000	0.000	5.000	0.000	0.000
Total	103.000	77.000	7.000	10.000	2.000	1.000	200.000	0.000	0.000
P_Accuracy	0.786	0.844	1.000	0.700	0.000	0.000	0.000	0.800	0.000
Kappa	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.667

ROC curve for SVM (radial basis)

ROC curve for SVM (polynomial)
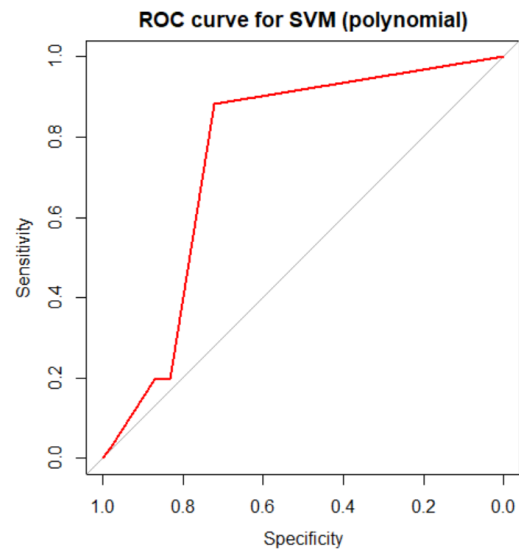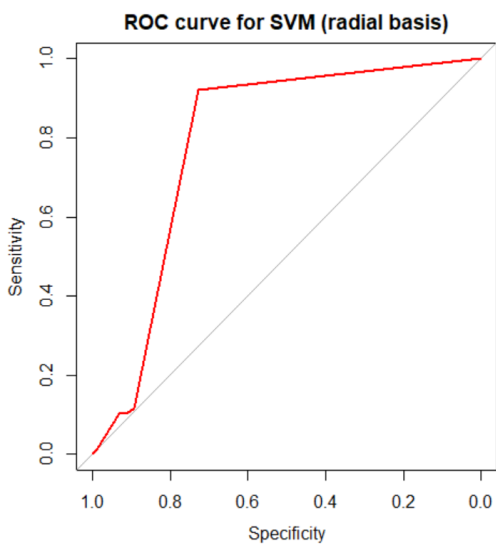
Maximum likelihood classifier (2011)									
Class	**1**	**2**	**3**	**4**	**5**	**6**	Total	U_Accuracy	Kappa
1	55.000	15.000	0.000	2.000	0.000	1.000	73.000	0.753	0.000
2	8.000	71.000	0.000	0.000	0.000	0.000	79.000	0.899	0.000
3	0.000	2.000	5.000	2.000	0.000	0.000	9.000	0.556	0.000

4	7.000	0.000	1.000	0.000	0.000	0.000	8.000	0.000	0.000
5	11.000	0.000	2.000	4.000	0.000	0.000	17.000	0.000	0.000
6	26.000	1.000	0.000	0.000	1.000	1.000	29.000	0.034	0.000
Total	107.000	89.000	8.000	8.000	1.000	2.000	215.000	0.000	0.000
P_Accuracy	0.514	0.798	0.625	0.000	0.000	0.500	0.000	0.614	0.000
Kappa	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.428

Maximum likelihood classifier (2015)									
Class	1	2	3	4	5	6	Total	U_Accuracy	Kappa
1	43.000	17.000	0.000	0.000	0.000	0.000	60.000	0.717	0.000
2	3.000	28.000	0.000	0.000	0.000	0.000	31.000	0.903	0.000
3	10.000	0.000	6.000	4.000	1.000	0.000	21.000	0.286	0.000
4	13.000	6.000	0.000	4.000	0.000	0.000	23.000	0.174	0.000
5	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000
6	34.000	26.000	0.000	2.000	1.000	1.000	64.000	0.016	0.000
Total	103.000	77.000	7.000	10.000	2.000	1.000	200.000	0.000	0.000
P_Accuracy	0.417	0.364	0.857	0.400	0.000	1.000	0.000	0.410	0.000
Kappa	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.238

R-codes

```
#Data Preparation----------------------------------------------------------------------
# Reading bands
band1 <- raster("b1_cl.tif")
band2 <- raster("b2_cl.tif")
band3 <- raster("b3_cl.tif")
band4 <- raster("b4_cl.tif")
band5 <- raster("b5_cl.tif")

#Preparation of the additional variables
NDVI1<- (band5 - band4)/(band5 + band4)
SAVI1<- ((band5 - band4)/(band5 + band4+0.5))*(1+0.5)
plot(SAVI1)
```

```
writeRaster(NDVI1, "NDVIL8", format = "GTiff", overwrite=TRUE)

writeRaster(SAVI1, "SAVIL8", format = "GTiff", overwrite=TRUE)

#Stitching of the images

image_stack <- stack(band1, band2, band3, band4, band5, NDVI1)

plot(image_stack)

#Training data

trainData <- readOGR("C:\\dimitrovgrad\\landsat8", layer="training_l8novi")

content_training8 <- summary(trainData)

content_training8

groundtruth <- readOGR("C:\\dimitrovgrad\\landsat8", layer = "points_accuracy15")

content_groundtruth <- summary(groundtruth)

content_groundtruth

#Classification process

setwd("C:\\dimitrovgrad\\landsat8")

library(sp)

library(raster)

library(rgdal)

library(rpart)

library(rasterVis)

library(caret)

library(randomForest)

library(snow)

library(e1071)

library(rgeos)

library(tiff)

library(raster)

library(RStoolbox)

library(caTools)
```

#Pre-processing

```
img<- brick("study_area_cl8.tif")

summary(img)

trainData <- readOGR("C:\\dimitrovgrad\\landsat8", layer="training_l8novi")

polygon <-SpatialPolygons(trainData)

summary(trainData)

plotRGB(img)

plotRGB(img, r=1, g=2, b=3,maxpixels=500000)

#Creation of the empty matrix

dfAll = data.frame(matrix(vector(), nrow = 0, ncol = length(names(img)) + 1))

responseCol <- "Class"
```

#Extract raster value based one training data

```
for (i in 1:length(unique(trainData[[responseCol]]))){

 category <- unique(trainData[[responseCol]])[i]

 categorymap <- trainData[trainData[[responseCol]] == category,]

 dataSet <- extract(img, categorymap)

 dataSet <- dataSet[!unlist(lapply(dataSet, is.null))]

 dataSet <- lapply(dataSet, function(x){cbind(x, class = as.numeric(rep(category,
nrow(x))))})

 df <- do.call("rbind", dataSet)

 dfAll <- rbind(dfAll, df)

}
```

```
#SVM radial basis

model.class <- svm(as.factor(class)~., data =dfAll,gamma=1, cost=1, type=NULL,
kernel="radial")

#Visualisation of results

pr <- predict(img, model.class,type='class',progress = 'text')

pr <- ratify(pr)

rat <- levels(pr)[[1]]
```

```
summary(rat)

rat$legent <- levels(category)

levels(pr) <- rat

levelplot(pr, maxpixels = 1e6, col.regions =
c("yellow","green","red","black","grey","blue"),

        scales=list(draw=TRUE),

        main = "Landsat 8 - SVM classification (radial basis)")

#export of results

writeRaster(pr, "SVMclassficationL8radial", format = "GTiff", overwrite=TRUE)


groundtruth <- readOGR("C:\\dimitrovgrad\\landsat8", layer = "svm_radialupdated15")

content_groundtruth <- summary(groundtruth)

content_groundtruth

#ROC Curve

library(pROC)

SVM1ROC <- as.numeric(groundtruth$Classified,groundtruth$GrndTruth, type =
"response")

summary(RFROC)

multiclass.roc(groundtruth$GrndTruth, SVM1ROC)

plotroc<- roc(groundtruth$GrndTruth, SVM1ROC)

plot.roc(groundtruth$GrndTruth, SVM1ROC, main = "ROC curve for SVM (radial
basis)", col = 'Red')

#Kappa statistics

library(asbio)

Kappa(groundtruth$RASTERVALU)
```

#SVM (linear)

```
model.class <- svm(as.factor(class)~., data =dfAll, cost=1, type=NULL, kernel="linear")

#Visualisation

pr1 <- predict(img, model.class,type='class',progress = 'text')

pr1 <- ratify(pr1)

rat <- levels(pr1)[[1]]
```

```
summary(rat)

rat$legent <- levels(category)

levels(pr1) <- rat

levelplot(pr1, maxpixels = 1e6, col.regions =
c("yellow","green","red","black","grey","blue"),

        scales=list(draw=TRUE),

        main = "Landsat 8 - SVM classification (linear)")

#export

writeRaster(pr1, "SVMclassficationL8l", format = "GTiff", overwrite=TRUE)

#Groundtruth_linear

groundtruth <- readOGR("C:\\dimitrovgrad\\landsat8", layer = "svm_linear_updated15")

content_groundtruth <- summary(groundtruth)

content_groundtruth
```

```
#ROC Curve

library(pROC)

SVM2ROC <- as.numeric(groundtruth$Classified,groundtruth$GrndTruth, type =
"response")

summary(RFROC)

multiclass.roc(groundtruth$GrndTruth, SVM2ROC)

plotroc<- roc(groundtruth$GrndTruth, SVM2ROC)

plot.roc(groundtruth$GrndTruth, SVM2ROC, main = "ROC curve for SVM (linear)",
col = 'Red')
```

#SVM polynomial

```
model.class <- svm(as.factor(class)~., data =dfAll,gamma=1, cost=1, type=NULL,
kernel="polynomial")

#Visualisation

pr2 <- predict(img, model.class,type='class',progress = 'text')

pr2 <- ratify(pr2)
```

```
rat <- levels(pr2)[[1]]

summary(rat)

rat$legent <- levels(category)

levels(pr2) <- rat

levelplot(pr2, maxpixels = 1e6, col.regions =
c("yellow","green","red","black","grey","blue"),

      scales=list(draw=TRUE),

      main = "Landsat 8 - SVM classification (polynomial)")

#export

writeRaster(pr2, "SVMclassficationL8pl", format = "GTiff", overwrite=TRUE)

#Groundtruth_linear

groundtruth <- readOGR("C:\\dimitrovgrad\\landsat8", layer = "svm_poly15updated")

content_groundtruth <- summary(groundtruth)

content_groundtruth




#ROC Curve

library(pROC)

SVM3ROC <- as.numeric(groundtruth$Classified,groundtruth$GrndTruth, type =
"response")

summary(RFROC)

multiclass.roc(groundtruth$GrndTruth, SVM3ROC)

plotroc<- roc(groundtruth$GrndTruth, SVM3ROC)

plot.roc(groundtruth$GrndTruth, SVM3ROC, main = "ROC curve for SVM
(polynomial)", col = 'Red')

#kappa statistics

library(asbio)
```

#Random forest classification

```
library(randomForest)

set.seed(1234)

# Random Search_tunning of the parameters
```

```
control <- trainControl(method="repeatedcv", number=10, repeats=3, search="random")
set.seed(1234)
mtry <- sqrt(ncol(x))
rf_random <- train(Class~., data=dataset, method="rf", metric=metric, tuneLength=15,
trControl=control)
print(rf_random)
plot(rf_random)


rf.md1 <- randomForest(as.factor(class)~., data =dfAll, ntree = 500, mtry = 3, proximity
= T, importance = T)

#Visualisation of results

pr1 <- predict(img, rf.md1,type='class',progress = 'text')

pr1 <- ratify(pr1)

rat <- levels(pr1)[[1]]

summary(rat)

rat$legent <- levels(category)

levels(pr1) <- rat

levelplot(pr1, maxpixels = 1e6, col.regions =
c("yellow","green","red","black","grey","blue"),

        scales=list(draw=TRUE),

        main = "Landsat 8 - RF classification")


writeRaster(pr1, "RandomforestL8", format = "GTiff", overwrite=TRUE)


#Groundtruth_linear

groundtruth <- readOGR("C:\\dimitrovgrad\\landsat8", layer = "svm_radialupdated15")

content_groundtruth <- summary(groundtruth)

content_groundtruth

#ROC Curve

library(pROC)

SVM4ROC <- as.numeric(groundtruth$Classified,groundtruth$GrndTruth, type =
"response")

summary(RFROC)
```

multiclass.roc(groundtruth$GrndTruth, SVM4ROC)

plotroc<- roc(groundtruth$GrndTruth, SVM4ROC)

plot.roc(groundtruth$GrndTruth, SVM4ROC, main = "ROC curve for Random Forest", col = 'Red')

#preparation of the data for ML classification

writeRaster(img, filename="image_stack.img", format="HFA", overwrite=TRUE)

#**Maximum Likelihood image classification**

classifyMLC <- superClass(image_stack,trainData, responseCol="Class",nSamples=154,model = "mlc", mode="classification",polygonBasedCV=TRUE, verbose=TRUE)

warnings()

model.mlc<- system.time(mlc <- predict(classifyMLC, img, filename="mlc_classification.img", na.rm=TRUE, overwrite = TRUE, inf.rm=FALSE, progress="text"))

#Visualisation of results

classifiedMLC <- raster("mlc_classification.img")

par(mar=c(3,3,2,2))

my_col <- c("blue", "grey", "black", "red", "green", "yellow")

image(classifiedMLC, main = "Maximum Likelihood", xlab = 700000, ylab = 480000, col = my_col)

2018

MONITORING AND EVALUATION OF THE BEST MAPPING APPROACH OF THE ENVIRONMENTAL IMPACTS AND LAND COVER CHANGES CAUSED WITH THE REVEAL OF THE QUARRY USING THE LANDSAT AND UAV IMAGES

Nikolina Nišić

Masters
Program
in **Geospatial**
Technologies