

# Klasifikacija litologije dijela sjevernog Velebita korištenjem multispektralnih satelitskih snimki Sentinel-2 u programskom jeziku R

---

Sziller, Laura

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering / Sveučilište u Zagrebu, Rudarsko-geološko-naftni fakultet**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:169:337317>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom](#).

Download date / Datum preuzimanja: **2024-07-24**



Repository / Repozitorij:

[Faculty of Mining, Geology and Petroleum Engineering Repository, University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU

RUDARSKO-GEOLOŠKO-NAFTNI FAKULTET

Diplomski studij Hidrogeologija i inženjerska geologija

**KLASIFIKACIJA LITOLOGIJE DIJELA SJEVERNOG VELEBITA  
KORIŠTENJEM MULTISPEKTRALNIH SATELITSKIH SNIMKI  
SENTINEL-2 U PROGRAMSKOM JEZIKU R**

Diplomski rad

Laura Sziller

GI 2120

Zagreb, 2023.



KLASA: 602-01/23-01/199  
URBROJ: 251-70-13-232  
U Zagrebu, 8.11.2023.

Laura Sziller, studentica

## RJEŠENJE O ODOBRENJU TEME

Na temelju vašeg zahtjeva primljenog pod KLASOM 602-01/23-01/199, URBROJ: 251-70-13-231 od 07.11.2023. priopćujemo vam temu diplomskog rada koja glasi:

### KLASIFIKACIJA LITOLOGIJE DIJELA SJEVERNOG VELEBITA KORIŠTENJEM MULTISPEKTRALNIH SATELITSKIH SNIMKI SENTINEL-2 U PROGRAMSKOM JEZIKU R

Za mentora ovog diplomskog rada imenuje se u smislu Pravilnika o izradi i obrani diplomskog rada Doc.dr.sc. Ivan Medved nastavnik Rudarsko-geološko-naftnog-fakulteta Sveučilišta u Zagrebu.


Mentor:

  
(potpis)

Doc.dr.sc. Ivan Medved

(titula, ime i prezime)


Predsjednica povjerenstva za  
završne i diplomske ispite:

  
(potpis)

Izv.prof.dr.sc. Ana Maričić

(titula, ime i prezime)

Prodekan za nastavu i studente:

  
(potpis)

Izv.prof.dr.sc. Borivoje  
Pašić

(titula, ime i prezime)

**KLASIFIKACIJA LITOLOGIJE DIJELA SJEVERNOG VELEBITA  
KORIŠTENJEM MULTISPEKTRALNIH SATELITSKIH SNIMKI SENTINEL-2 U  
PROGRAMSKOM JEZIKU R**

Laura Sziller

Rad izrađen: Sveučilište u Zagrebu  
Rudarsko-geološko-naftni fakultet  
Zavod za geofizička istraživanja i rudarska mjerenja  
Pierottijeva 6, 10 000 Zagreb

Sažetak

Svrha ovog rada je klasifikacija litologije određenog područja korištenjem multispektralnih satelitskih snimaka i programa R, izradio se algoritam koji se može primijeniti za bilo koje područje istraživanja i klasifikaciju litologije područja. Korištenjem softvera QGIS, izrađeni su poligoni starosti i vrste stijena kao ulazni podaci za napravljeni algoritam. Prostorna analiza se dobila pomoću R-a, preko integriranog razvojnog okruženja RStudio-a. Korištenjem nadzirane klasifikacije sustav se uvijekbao, dobio je podatke te na temelju stečenog znanja detektirao nove podatke povlačenjem sličnosti između njih. Rezultat istraživanja je grafički prikaz tri modela rf, svm i nnet, gdje se odabire pouzdaniji model koji sadrži litologiju s ostalim klasama i spektralni potpisi multispektralnih satelitskih snimki Sentinel-2. Rad se može upotrijebiti kao podloga budućim istraživanjima u svrhu pronalaska naprednijeg sustava detekcije litologije područja istraživanja, koja sadržava informacije o starosti i vrsti stijena, čime inženjerima geologije omogućava brže i ekonomičnije istraživanje i razumijevanje geoloških karakteristika terena.

Ključne riječi: Klasifikacija, litologija, Sentinel-2, R, RStudio, QGIS, RF, SVM, NNET, vrsta stijene, starost stijene, nadzirana klasifikacija, spektralni potpis

Diplomski rad sadrži: 56 stranice, 9 tablice, 33 slike, 0 priloga i 43 referenca.

Jezik izvornika: Hrvatski

Pohrana rada: Knjižnica Rudarsko-geološko-naftnog fakulteta, Pierottijeva 6, Zagreb

Mentor: Dr. sc. Ivan Medved, docent RGNF

Ocjenjivači: Dr. sc. Ivan Medved, docent RGNF  
Dr. sc. Jasna Orešković, izvanredni profesor RGNF  
Dr. sc. Tomislav Korman, izvanredni profesor RGNF

**CLASSIFICATION OF LITHOLOGY IN THE NORTHERN VELEBIT USING MULTISPECTRAL  
SENTINEL-2 SATELLITE IMAGERY IN THE R PROGRAMMING LANGUAGE**

Laura Sziller

Thesis completed at: University of Zagreb

Faculty of Mining, Geology and Petroleum Engineering  
Institute of Geophysical Exploration and Mine Surveying  
Pierottijeva 6, 10 000 Zagreb

Abstract

The purpose of this study is the classification of lithology in a specific area using multispectral satellite imagery and the R programming language. An algorithm has been developed that can be applied to any research area for lithology classification. Using QGIS software, polygons of age and rock types were created as input data for the developed algorithm. Spatial analysis was performed using R through the integrated development environment of RStudio. Through supervised classification, the system was trained to acquire data and, based on the acquired knowledge, to detect new data by drawing similarities between them. The research results are graphically presented through three models: random forest (rf), support vector machine (svm), and neural network (nnet). The study aims to select the most reliable model that incorporates lithology with other classes and spectral signatures from Sentinel-2 multispectral satellite imagery. This work can serve as a foundation for future research aimed at developing an advanced system for detecting the lithology of research areas, including information about age and rock type. Such a system would enable geology engineers to conduct faster and more cost-effective research, enhancing their understanding of geological terrain characteristics.

Keywords: Classification, lithology, Sentinel-2, R, RStudio, QGIS, RF, SVM, NNET, rock type, rock age, supervised classification, spectral signature

Thesis contains: 56 pages, 9 tables, 33 figures, 0 appendixes and 43 references.

Original in: Croatian

Archived in: Library of Faculty of Mining, Geology and Petroleum Engineering, Pierottijeva 6, Zagreb

Supervisors: PhD. Ivan Medved, Assistant Professor RGNF

Reviewers: PhD Ivan Medved, Assistant Professor RGNF  
PhD Jasna Orešković, Assistant Professor RGNF  
PhD Tomislav Korman, Assistant Professor RGNF

## **Popis i objašnjenje kratica korištenih u istraživanju**

**AP** - Amfibolski porfirit

**ArcGIS** - Geografski informacijski sustav (GIS) softver razvijen od strane tvrtke Esri

**B** - Breče

**BR** – Omjer pojasa

**BV** - Brečoliki vapnenci

**CK** - Crveni klasititi

**D** – Dolomiti

**DV** - Dolomitizirani vapnenci

**FCC** – Kompozit lažnih boja

**FV** - Foraminiferski vapnenci

**GADM** - Global Administrative Areas

**GIS** - Geografski informacijski sustav

**GK** – Gornja kreda

**GS** - Glaciofluvijalni sedimenti

**GT** - Gornji trijas

**IDE** - Integrated development environment

**JD** – Jura doger

**JL** – Jura lijas

**JM** - Jura malm

**KIK** - Konglomerati i klastiti

**KV** - Kvaratar

**LiDAR** - Light Detection and Ranging

**MLC** - Klasifikator maksimalne vjerojatnosti

**MNF** - Minimalni udio šuma

**NNET** – Neuronska mreža (Neural Network)

**OLI** - Operation Land Imager

**OGK** - Osnovna geološka karta

**P** - Pijesci

**PA** - Paleogen

**PCA** - Glavnica analize komponenti

**PIS** - Pijesak i šljunak

**PN** – Paleogen-neogen

**QGIS** – Geografski informacijski sustav otvorenog koda (*Quantum Geographic Information System*)

**R** – Programski jezik

**RF** – Slučajna šuma (*Random Forest*)

**RGB** - Red, green and blue

**RS** - Remote sensing

**RV** - Rudistni vapnenci

**S2A/B** - Sentinel-2A i Sentinel-2B

**S2MSI2A** - MultiSpectral Instrument (MSI) na Sentinel-2A

**SFRJ** - Socijalistička Federativna Republika Jugoslavija

**ST** - Srednji trijas

**SVM** - Support Vector Machine

**TIFF** – Format datoteke za pohranu slika (*Tagged Image File Format*)

**TIRS** - Thermal Infrared Sensor

**V** – Vapnenci

**VID** - Vapnenci i dolomiti

**WFS** – Mrežni servis za vektorske podatke (*Web Feature Service*)

**WMS-C** - Web Map Service-Cached

**WMS** – Mrežni kartografski servis (*Web Map Service*)

**WMTS** - Web Map Tile Service



## Sadržaj

1. Uvod.....	1
2. Ciljevi rada.....	2
3. Područje istraživanja.....	3
3.1. Prethodna istraživanja.....	6
4. Materijali i metode.....	8
4.1. Copernicus (ESA) i Sentinel-2.....	8
4.2. QGIS .....	10
4.3. R programski jezik.....	11
4.3.1 R paketi .....	12
4.4. Metode strojnog učenja za klasifikaciju.....	14
5. Analiza spektralnog potpisa satelitske misije Sentinel-2 .....	16
6. Obrada podatka metodama strojnog učenja .....	19
6.1. Korištenje metode <i>randomForest</i> (RF) iz paketa CARET.....	21
6.2. Korištenje metode <i>Support Vector Machine</i> (SVM) iz paketa CARET .....	25
6.3. Korištenje metode <i>Neural Network</i> (NNET) iz paketa CARET .....	27
6.4. Usporedba dobivenih modela.....	30
7. Rezultati i rasprava korištenjem Sentinel-2 .....	32
7.1. Potvrda modela rf starosti stijena sa kartama OGK Rab i Otočac .....	37
8. Zaključak.....	42
9. Zahvale.....	43
10. Literatura.....	44

Popis slika:

Slika 2.1. Prikaz rada algoritma.....	2
Slika 3.1. Područje istraživanja općina Senj, otok Rab, otok Sveti Grgur i Goli otok.....	3
Slika 4.3.1 Granice područja istraživanja na OGK Rab i OGK Otočac.....	4
Slika 4.1.1. Copernicus Sentinel-2 (Sentinel Online, 2023.).....	8
Slika 4.1.2. Copernicus – dobivanje i preuzimanje podataka (Copernicus-b, 2023.).....	9
Slika 4.2.1. Područja istraživanja programu QGIS.....	11
Slika 4.3.1. RStudio.....	12
Slika 4.4.1. Prikaz rada RF-a (Javatpoint-b, 2021.).....	14
Slika 4.4.2. Prikaz rada SVM-a (Datacamp, 2018.).....	15
Slika 4.4.1. Prikaz rada Neural Network (Geeksforgeeks, 2020.).....	15
Slika 5.1. Spektralni potpis starosti stijena Sentinel-2.....	17
Slika 5.2. Spektralni potpis vrste stijena Sentinel-2.....	18
Slika 6.1. Prikaz histograma multispektralnih snimaka.....	20
Slika 6.2. Grafički prikaz predikcije RF-a.....	22
Slika 6.3. Prikaz <i>heatmap1</i> .....	24
Slika 6.4. Prikaz <i>heatmap2</i> .....	25
Slika 6.2.1. Grafički prikaz modela SVM-a.....	26
Slika 6.3.1. Grafički prikaz modela <i>Neural Network</i> .....	28
Slika 6.3.2. Prikaz neuronske mreže.....	29
Slika 6.3.3. Dijagram neuronske mreže.....	30
Slika 6.4.1. Prikaz rezultata unakrsne validacije.....	31
Slika 7.1. Predict_rf starost stijena.....	32
Slika 7.2. Predict_rf vrste stijena.....	33
Slika 7.3. Predict_svm starost stijena.....	34
Slika 7.4. Predict_svm vrste stijena.....	34
Slika 7.5. Predict_nnet starost stijena.....	35
Slika 7.6. Predict_nnet vrste stijena.....	36

Slika 7.1.1. Područje istraživanja na OGK Rabu i OGK Otočcu.....	37
Slika 7.1.2. Klasifikacija litologije modela <i>rf</i> . .....	38
Slika 7.1.3. Usporedba litologije modela <i>rf</i> i OGK Otočac. ....	39
Slika 7.1.4. Razlika litologije modela <i>rf</i> naspram OGK Otočac.....	39
Slika 7.1.5. Usporedba litologije modela <i>rf</i> i OGK Rab. ....	40
Slika 7.1.6. Razlika litologije modela <i>rf</i> naspram OGK Rab.....	41

Popis tablica:

Tablica 3.1. Starost i vrsta stijena istraživnog područja.....	5
Tablica 4.1.1. Sentinel-2 kanali.....	10
Tablica 4.3.1.1. Tablica paketa u R-u.....	13
Tablica 6.1. Prikaz vrijednosti dobivene tablice. ....	19
Tablica 6.2. Količina treninga podataka klasa. ....	20
Tablica 6.3. Količina testa podataka klasa. ....	21
Tablica 6.1.1. Statistički prikaz <i>random forest</i> klasifikacije.....	23
Tablica 6.2.1. Statistički prikaz <i>svm</i> klasifikacije. ....	26
Tablica 6.3.1. Statistički prikaz <i>nnet</i> klasifikacije.....	28

## 1. Uvod

Geološka klasifikacija litologije predstavlja ključni aspekt geoloških istraživanja i razumijevanja površine Zemlje i procesa koji su oblikovali njezinu geološku povijest. Poznavanje informacija o starosti i vrsti stijena na određenom geografskom području temelj je za bolje upravljanje prirodnim resursima, planiranje infrastrukturnih projekata, razumijevanje geoloških procesa i karakteristika terena na globalnoj razini, procjene rizika od prirodnih katastrofa, očuvanje okoliša i brojne sektore, uključujući geološku inženjerstvo, rudarstvo... Tradicionalne metode geološkog kartiranja i terenskog istraživanja iako dragocjene, često su vremenski i resursno zahtjevne, a rezultati su ograničeni terenskim pristupom i podložne ljudskoj subjektivnosti. Razvoj tehnologije satelitskih snimaka i napredak u analizi podataka otvara nove mogućnosti za potrebe istraživanja i dobivanja učinkovitijih rezultata, što rezultira većom efikasnošću i smanjenjem troškova u usporedbi s tradicionalnim metodama terenskog istraživanja.

S obzirom na spomenuto, glavni cilj ovog rada je klasifikacija litologije određenog područja korištenjem multispektralnih satelitskih snimaka i programa R. Unutar ovog rada prikupljeni su podaci sa multispektralnih satelitskih snimaka dobivenih pomoću Sentinel-2 satelita. Korištenjem softvera QGIS, izrađeni su poligoni koji će biti ulazni podaci za napravljeni algoritam. Prostorna analiza dobiva se pomoću R-a, programskog jezika za statističko računanje i grafiku, preko integriranog razvojnog okruženja RStudio-a. Korištenjem nadzirane klasifikacije sustav se prvo mora uvježbati. Sustav dobiva podatke te na temelju stečenog znanja može detektirati nove podatke povlačenjem sličnosti između njih. U klasifikaciji, program koristi skup podataka da nađe nove te ih kategorizira u različite klase ili grupe. Uz pomoć novo nastalog algoritma dobiva se klasifikacija litologije područja istraživanja koja sadržava informacije o starosti i vrsti stijena, čime inženjerima geologije omogućava brže i ekonomičnije istraživanje i razumijevanje geoloških karakteristika terena.

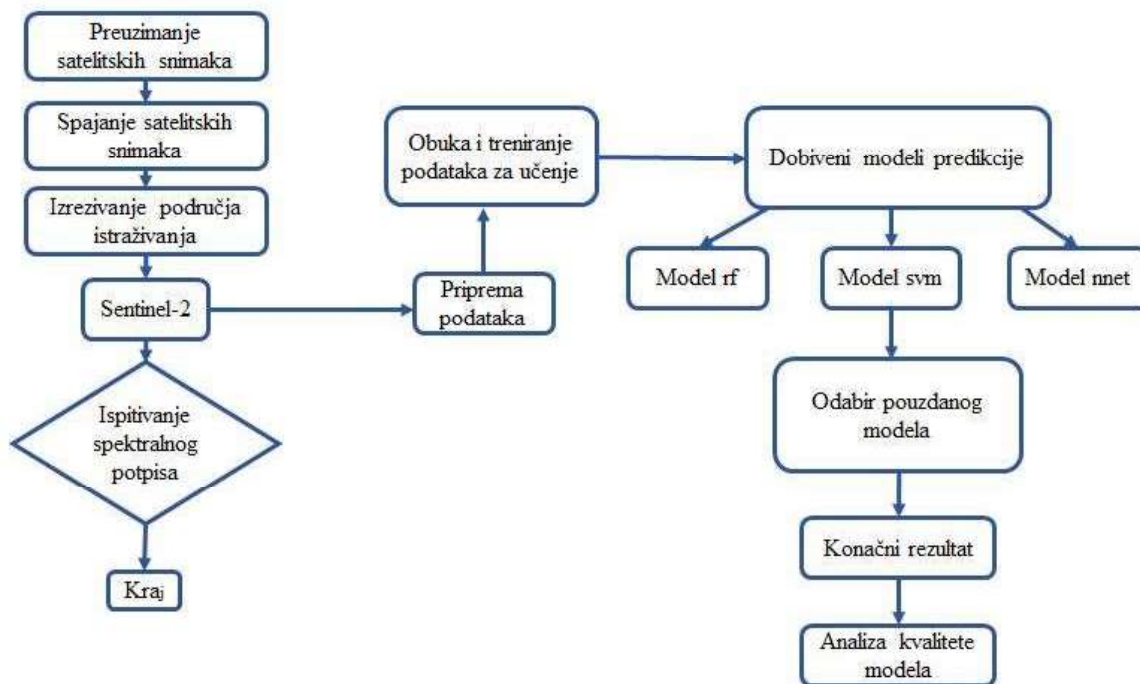
Ovaj rad prikazuje primjenu programskog jezika R za potrebe istraživanja litologije u Republici Hrvatskoj što je prikazano kroz rezultate rada.

## 2. Ciljevi rada

Cilj ovog rada bio je napraviti algoritam u programskom jeziku R pomoću kojeg se može klasificirati litologija, koja će sadržavati informacije o starosti i vrsti stijena određenog terena.

Za potrebe ovog rada prikupljaju se podaci sa multispektralnih satelitskih snimaka pomoću europske svemirske agencije Copernicus (Sentinel-2). Uzimaju se multispektralne satelitske snimke područja Republike Hrvatske na kojima će se provesti istraživanje. Na temelju Osnovne geološke karte SFRJ Otočac, 1: 100 000 i Osnovne geološke karte SFRJ Rab, 1: 100 000, primjenjujući program QGIS izrađuju se dvije grupe poligona. Jedna grupa poligona radi se za vremensko razdoblje stijena, a druga grupa poligona za vrstu stijena, koje će služiti kao ulazni podaci za uvježbavanje sustava. U programskom jeziku R kreira se kod koji predstavlja nadziranu klasifikaciju koja na temelju sličnosti danih poligona pronalazi nove lokacije.

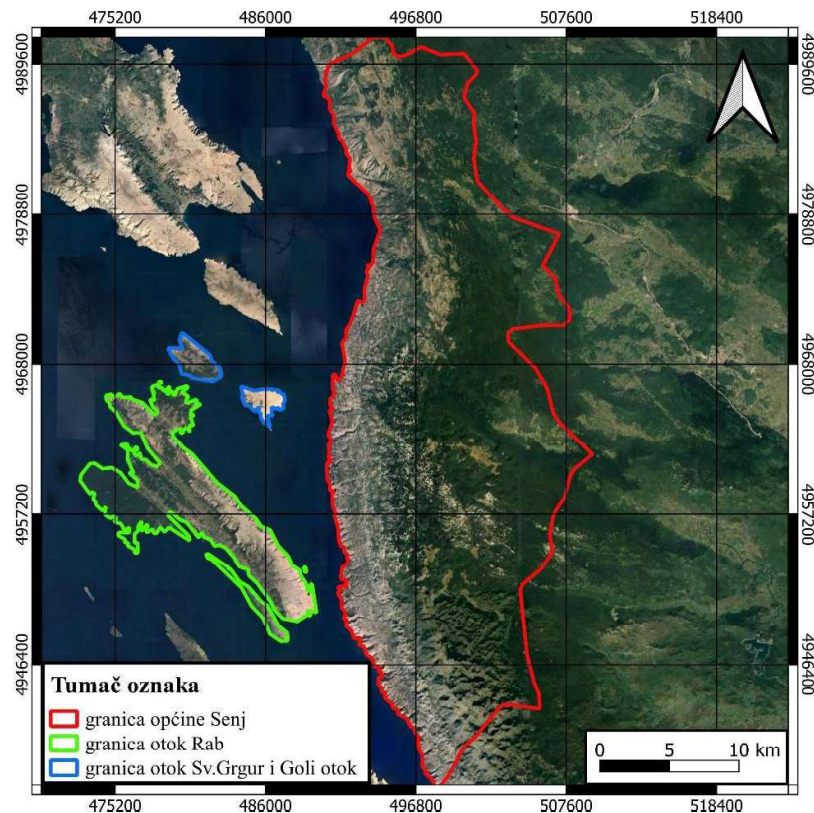
Glavni cilj je dobiti algoritam koji se može primijeniti za bilo koje područje istraživanja i klasifikaciju litologije u svrhu geološkog kartiranja za bolje razumijevanje geoloških procesa i karakteristika terena, rudarstva, urbanističkog planiranja, inženjerstva i zaštite okoliša. Tijek izrade procesa algoritma prikazan je na slici 2.1.



Slika 2.1. Prikaz rada algoritma.

### 3. Područje istraživanja

Za područje istraživanja izdvojeni su općina Senj, otok Rab, otok Sveti Grgur i Goli otok. Osnovne informacije o geološkoj starosti i vrsti stijena, ključne su za razumijevanje geoloških karakteristika svakog područja i geološke povijesti za daljnju analizu klasifikacije litologije. Na kartama izrađenih u programu QGIS prikazane su granice istraživanog područja (Slika 3.1.) i (Slika 3.2.).

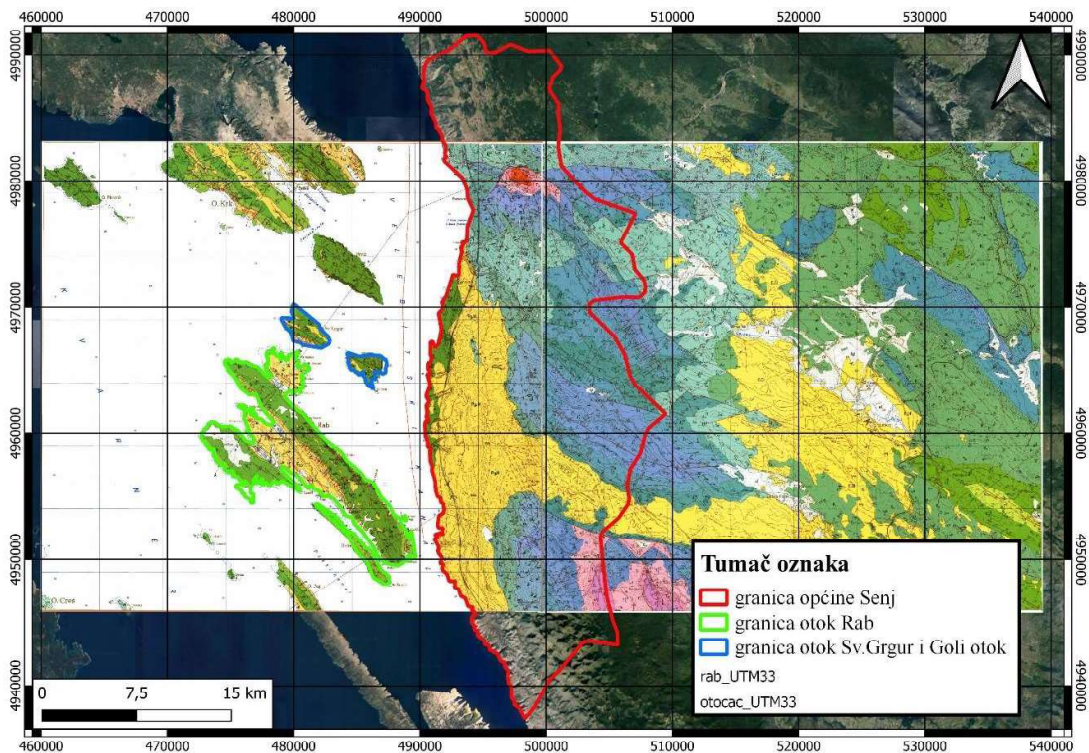


Slika 3.1. Područje istraživanja općina Senj, otok Rab, otok Sveti Grgur i Goli otok.

Općina Senj, na sjevernom Jadranu, smještena duž obale Jadranskog mora i okružena planinom Velebit, zanimljiva je svojom geološkom raznolikošću. Prema starosti u široj okolini Senja obuhvaćenoj geološkom kartom, razlikuju se trijasko vapnenačke, dolomitne, klastične i eruptivne stijene, jurske, kredne i tercijarne karbonatne naslage i kvartarne tvorevine. Naslage koje izgrađuju senjsku okolicu odnosno primorske padine sjevernog Velebita i Velike Kapele nastale su u uvjetima plitkomorske sedimentacije. U svakom vremenskom razdoblju, osim toga, karbonatske naslage karakteriziraju određene zajednice fosila (Milan, 1970.).

Otok Rab, smješten u sjevernom dijelu Jadranskog mora, poznat po svojoj prekrasnoj obali, skriva u sebi bogatu i složenu geološku povijest. Izgrađen od krednih, paleogenskih i kvartarnih naslaga. Najveću rasprostranjenost imaju naslage krede i paleogena, dok su naslage kvartara u odnosu na prethodne, znatno slabije razvijene (Mamužić, 1962.).

Otok Sveti Grgur i Goli otok, poznati po svojoj povijesti kao zatvorski otoci, također posjeduju svoje jedinstvene geološke karakteristike. Imaju strmu konfiguraciju terena i prevladavaju vapnenačke stijene. Otok Sveti Grgur, na Jadranu, najsjeverniji zimzeleni otok mediteranskog izgleda, nalazi se između otoka Raba i Krka. Goli otok je smješten između sjeveroistočnog dijela otoka Raba i kopnene obale u sjevernom dijelu Velebitskog kanala (MedMPAnet projekt, 2014.).



Slika 4.3.1 Granice područja istraživanja na OGK Rab i OGK Otočac.

Starost i vrsta stijena koje se nalaze na području istraživanja prikazani su u tablici 3.1. Na području otoka Raba, otoka Sv. Grgura i Golog otoka dominira geološko razdoblje krede, na ovim otocima također se javljaju stijene iz razdoblja paleogena i kvartara. Općina Senj, svojom raznolikošću geoloških karakteristika, nudi širok spektar geoloških razdoblja. Glavnim dijelom prevladava geološko razdoblje jure, zatim paleogena. U manjem dijelu prisutni su i ostali ključni geološki periodi, razdoblja krede, trijasa i kvartara također obogaćuju geološki mozaik ovog područja.



Tablica 3.1. Starost i vrsta stijena istraživanog područja

Starost	Geološka oznaka	Vrsta stijene
Kvartar	p	Pijesci
Kvartar	d	Pijesak i šljunak
Kvartar	fgl	Glaciofluvijalni sedimenti
Paleogen, neogen	Pg, N	Breče
Paleogen	$E_{2,3}$	Lapori i pješčenjaci
Paleogen	$E_2^2$	Lapori
Paleogen	$E_{1,2}$	Foraminiferski vapnenci
Gornja kreda	$K_2^{2,3}$	Rudistni vapnenci
Gornja kreda	$K_2^{1,2}$	Vapnenci
Gornja jura, malm	$J_3^{2,3}$	Vapnenci
Gornja jura, malm	$J_3^{2,2}$	Vapnenci
Gornja jura, malm	$J_3^{1,2}$	Vapnenci i dolomiti
Srednja jura, doger	$J_2$	Brečoliki vapnenci
Donja jura, lijas	$J_1^4$	Dolomitizirani vapnenci
Donja jura, lijas	$J_1^3$	Vapnenci
Gornji trijas	$T_3^{2,3}$	Dolomiti
Gornji trijas	$T_3^{1,2}$	Konglomerati i klastiti
Gornji trijas	$T_3^1$	Crveni klasititi
Gornji trijas	$T_3$	Dolomiti
Srednji trijas	$T_2^2$	Vapnenci i dolomiti
Srednji trijas	$T_2$	Vapnenci
Srednji trijas	a	Amfibolski porfirit

### 3.1. Prethodna istraživanja

Pregledom dostupne znanstvene literature iz područja daljinskih istraživanja u praćenju pojave litologije izdvajaju se sljedeći rezultati.

Primjena daljinskog istraživanja za područje Črni Kal, Slovenija. Korištenje satelitskih snimaka u inženjerskogeološkom kartiranju većih mjerila. Pokušaj utvrđivanja daljinskog očitavanja (RS) za lociranje faktora kao što su litologija i prisutnost rasjeda. Na području Kal provedena je digitalna obrada slike. Korištene su dvije različite metode klasifikacije, jedna nenadzirana klasifikacija za kompozit boja traka 1, 2 i 7, i nadzirana klasifikacija za sastav boja od traka 2, 4 i 6. Nakon klasifikacije izračunata je točnost za obje metode klasifikacije (Komac et al., 1998.). Upotrebom satelita Landsat 8 moguće je praćenje sjevernih ekosustava Bovanenkove regije za otkrivanje promjene okoliša prouzročene raznim čimbenicima. Istraživanja se mogu unaprijediti korištenjem daljinskih istraživanja i softvera GIS za potrebe kartiranja i analize prostorno-vremenskih promjena u vegetacijskim vrstama i pokrovu (Lemenko et al., 2012.).

Geomorfološko modeliranje reljefa terena Italije i regionalno kartiranje korištenjem programskog jezika R. Geomorfološko modeliranje reljefa terena za dobivanje nagiba, aspekta, sjene, nadmorske visine i njihove vizualizacije dobiveno je upotrebom R paketa, raster i tmap. Regionalno kartiranje Italije dobiva se pomoću R paketa ggmap i ggplot2 (Lemenkova, 2020.).

Dobrinić i Gašparović koriste Sentinel-1 i Sentinel-2 snimke za kartiranje vegetacije, korištenjem slučajne klasifikacije šuma (RF) za područje sjeverne Hrvatske. RandomForest su odabrali zbog mogućnosti jednostavnih parametara, procjena važnosti i brzog vremena izračuna u klasifikaciji. Različite klasifikacije i procjene točnosti programa izvršene su pomoću programskog jezika R i njegovog sučelja RStudio-a. Najveću točnost u istraživanju, raspona od 89% do 92% za kartiranje pokrova zemljišta, daju RF i SVM metode. Zbog sposobnosti rješavanja visokodimenzionalnih podataka (Dobrinić et al., 2021.). Razvoj algoritma klasifikacije za otkrivanje plutajuće plastike pomoću daljinskih istraživanja korištenjem Sentinel-2 snimke. Zbog široke razlučivosti na daljinu upotrebom Sentinel-2 snimaka moguće je precizno otkriti plastični otpad za potrebe sanacije plastike iz oceana (Basu et al., 2021.). Istraživanje problema klasifikacije urbanog pokrova zemljišta u Peking (Kina), korištenjem satelita S2A/B za daljinska istraživanja. Optimizacija RF-a za usporedbu s ostalim metodama za klasifikaciju (5 klasa). Uspoređuje se optimizirani RF Bayesove optimizacije sa SVM-om i RF-om s danim hiperparametrima (Zhang et al., 2021.).

Na području Nacionalnog Parka Sjeveri Velebit provedena su se daljinska istraživanja pomoću multispektralne analize Landsat 8 snimke. Radi se o krškom području u kojem se uglavnom izmjenjuju kalcitne stijene sa kalcitnim stijenama s dolomitima. Pomoću Landsat 8 snimke i kombinacijama

raznih omjera spektralnih kanala s kojima je napravljen FCC-a (false color composite) izdvojene su litološke jedinice na području Sjevernog Velebita. Usporedbom litoloških jedinica dobivenih iz FCC-a i litoloških jedinica iz OGK-a dobiven je okvir za procjenu preciznosti pri izvlačenju litoloških jedinica pomoću multispektralne analize satelitskih snimaka (Jukić, 2021.).

Provedena su daljinska istraživanja, LiDAR i RGB snimanja, terenska i laboratorijska istraživanja na istraživačkom području Galgovo. Prikupljeni podaci analizirani su i prikazani pomoću softvera ArcGIS. Podaci dobiveni daljinskim istraživanjem korišteni su za izradu inventara klizišta i karata gustoće klizišta, energije reljefa i nagiba. Terenskim i laboratorijskim istraživanjima dobiveni su podaci o granulometriji i litologiji istraživanog područja (Breković, 2022.).

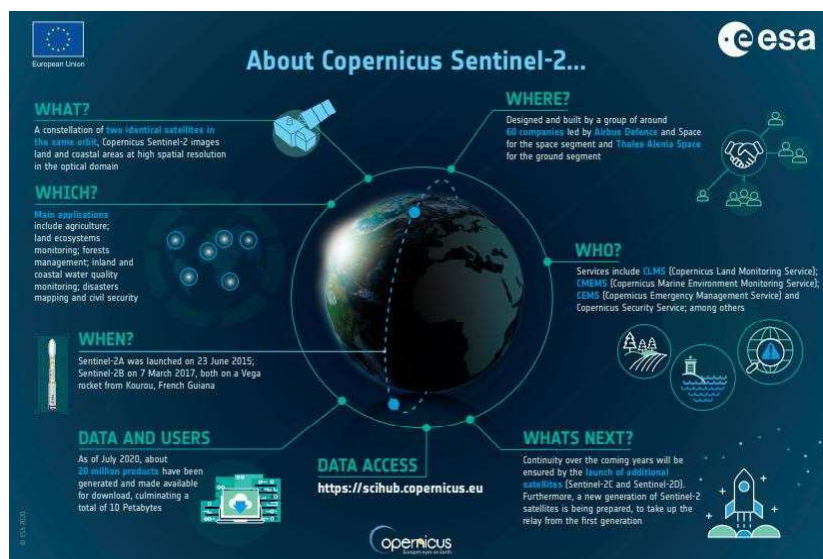
Upotreba višesenzorskih satelitskih podataka i terenskih studija za otkrivanje strukturne evolucije i metalogenije zlata Gerfa, Ofiolitski pokrivač, Istočna pustinja, Egipat. Daljinski se integriraju multisenzorski podaci, strukturne analize i istraživanje terena u otkrivanju geoloških i strukturnih kontrola mineralizacije zlata u području Gabal Gerf. Multispektralni optički senzori Landsat-8 OLI/TIRS (L8) i Sentinel-2B (S2B) obrađeni su za mapiranje jedinica litoloških stijena. Algoritmi za obradu slike uključujući kompozit lažnih boja (FCC), omjer pojasa (BR), glavnicu analiza komponenti (PCA), minimalni udio šuma (MNF) i klasifikator maksimalne vjerojatnosti (MLC) bili su učinkoviti u izradi sveobuhvatne geološke karte područja. Informacije izvučene iz podataka daljinski očitanih, strukturna interpretacija i terenski rad korišten je za izradu karte potencijalnih zona mineralizacije zlata koju je potvrdio referentna i terenska opažanja. Ova studija pokazuje mogućnosti daljinskog očitavanja za identifikaciju alteracijskih zona i strukturne kontrole mineralizacije zlata u visokim deformirane ofiolitičke regije (Abd El-Wahed et al., 2023.).

## 4. Materijali i metode

U svrhu istraživanja ovog rada prikazat će se postupci metoda za prikupljanje i obradu podataka. Za prikupljanje podataka sa multispektralnih satelitskih snimaka istraživanog područja korištene je Copernicus (Sentinel-2) satelitska misija. Obrada prikupljenih podataka napravljena je pomoću programa QGIS i R programskog jezika. Prednost korištenih materijala i metoda je ta što su besplatni i lako dostupni.

### 4.1. Copernicus (ESA) i Sentinel-2

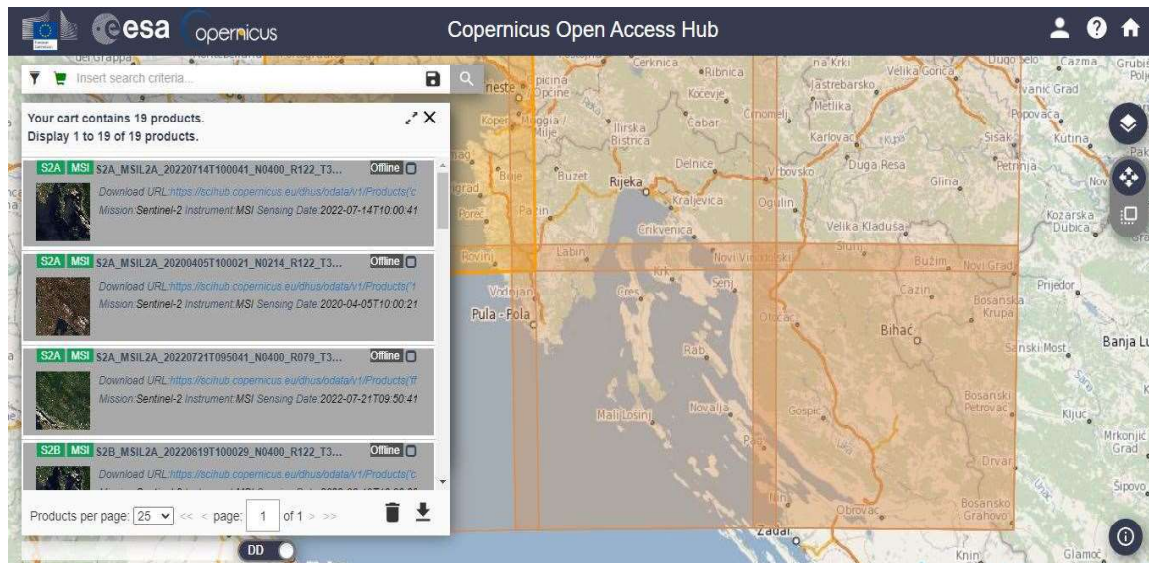
Copernicus je program Europske unije za promatranje Zemlje (Slika 4.1.1.). Copernicus usluge se mogu upotrebljavati u različitim područjima kao što su urbana područja, zaštita prirode, poljoprivreda, šumarstvo, ribarstvo, zdravstvo, infrastrukturu, turizam...(Copernicus-a, 2023.). Program Copernicus je besplatan i dostupan na web stranici Copernicus Open Access Hub (Slika 4.1.2.).



Slika 4.1.1. Copernicus Sentinel-2 (Sentinel Online, 2023.).

Sentinel-2 je europska misija za snimanje širokog pojasa, visoke razlučivosti i više spektralnih slika. Posjeduje teret optičkog instrumenta koji stvara 13 spektralnih kanala, četiri kanala na 10m, šest kanala na 20m i tri kanala na 60m prostorne rezolucije. Misija Sentinel-2 sastoji se od dva satelita blizanca, Santinel-2A i Sentinel-2B. dva satelita rade istovremeno, pod kutom od 180°, jedan prema drugome u orbiti sinkronizirano sa Suncem na srednjoj visini od 786km. Cilj misije Sentinel-2 daje značajan doprinos programa Copernicus za klimatske promjene, nadzor zemljišta, upravljanje hitnim slučajevima i sigurnosti. Omogućava kvalitetne podatke za klasifikaciju zemljišnog pokrova, atmosfersku korekciju i smanjivanje raspona oblaka ili snijega (Sentinel Online, 2023.).

Odabirom poligona područja, vremenskog perioda, postotka oblaka i produkta S2MSI2A preuzima se multispektralna satelitska snimka (Copernicus-b, 2023.). Sustav generira multispektralne satelitske snimke koje odgovaraju vremenskom periodu i traženom području. Dobivene snimke su 10m, 20m i 60m prostorne rezolucije.



Slika 4.1.2. Copernicus – dobivanje i preuzimanje podataka (Copernicus-b, 2023.).

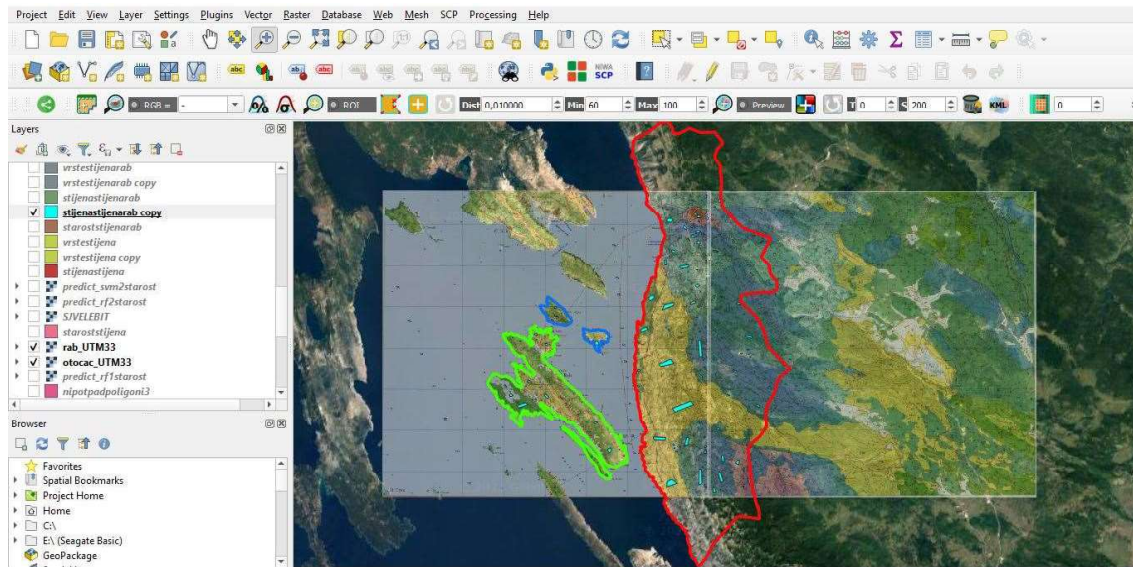
Kao što je navedeno prije multispektralna satelitska snimka se sastoji od 13 kanala kod koji svaki može varirati između 10m, 20m i 60 m rezolucije piksela. Rezoluciju od 10m imaju kanali B2 (plavi), B3 (zeleni), B4 (crveni) i B8 (bliski infracrveni). Kanali s crvenom (B5), bliskim infracrvenom NIR bojom (B6, B7, B8a) i kratkovalni infracrveni SWIR (B11, B12) imaju 20 metarsku rezolucije. Od 60 metarske rezolucije ostaju kanali B1 (obalna aerosol) i B10 (infracrveni SWIR). Kombinacijom kanala mogu se bolje isticati određene značajke na multispektralnoj satelitskoj snimci, a to mogu biti geološke, poljoprivredne ili vegetacijske. U ovome radu koristi se kombinacija 10 kanala, B2 (plava), B3 (zeleni), B4 (crveni), B5 (VNIR), B6 (VNIR), B7 (VNIR), B8 (VNIR), B8a (VNIR), B11 (SWIR), B12 (SWIR) i njihovih boja kojima se prikazuje slika na isti način kako ljudsko oko vidi svijet. Vegetacija je zelena, urbana područja su bijela i siva, a područje s vodom je uglavnom tamnoplave boje ovisno o tome koliko je čisto (GISGeography, 2022.).

Tablica 4.1.1. Sentinel-2 kanali.

Kanali	Rezolucija (m)	Valna duljina (nm)	Opis
B1	60	443	Ultra plava
B2	10	490	Plava
B3	10	560	Zelena
B4	10	665	Crvena
B5	20	705	Vidljivi i bliski infracrveni (VNIR)
B6	20	740	Vidljivi i bliski infracrveni (VNIR)
B7	20	783	Vidljivi i bliski infracrveni (VNIR)
B8	10	842	Vidljivi i bliski infracrveni (VNIR)
B8a	20	865	Vidljivi i bliski infracrveni (VNIR)
B9	60	940	Kratkovalno infracrveni (SWIR)
B10	60	1357	Kratkovalno infracrveni (SWIR)
B11	20	1610	Kratkovalno infracrveni (SWIR)
B12	20	2190	Kratkovalno infracrveni (SWIR)

## 4.2. QGIS

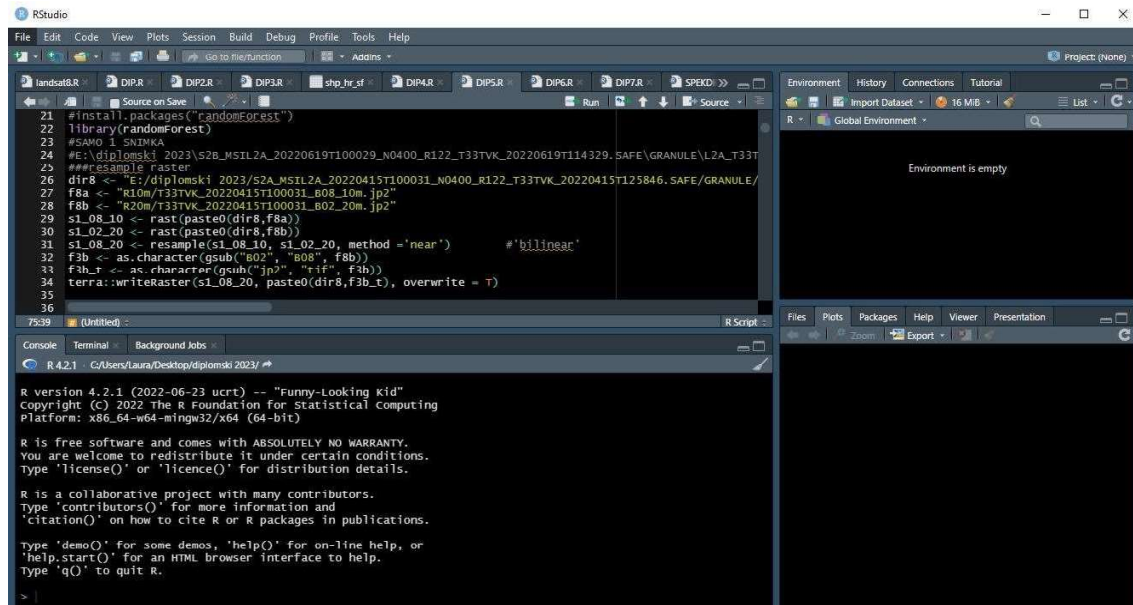
QGIS je geografski informacijski sustav otvorenog koda s besplatnom upotrebom koji podržava vektorske i rasterske podatke te baze podataka. U QGIS-u je moguća nadogradnja mnogih GIS funkcija sa osnovnim značajkama. Mogu se pregledati, uređivati, stvarati, upravljati i izvoziti vektorski i rasterski podaci u različitim formama, kao na primjer karte. One koje su objavljene na internetu mogu se pretraživati i unositi u QGIS sustav kao WMS, WMTS, WMS-C, WFS itd. Instalacijom UMN MapServer-a ili GeoServer-a s webserverom je moguće objaviti podatke na internetu (QGIS-b, 2023., QGIS-a, 2023.). U programu QGIS korištenjem multispektralnih satelitskih snimaka određene su granice istraživanog područja te se izrađuju se dvije grupe poligona, za vremensko razdoblje i vrstu stijena, koje će služiti za daljnju analizu klasifikacije litologije. (Slika 4.2.1.).



Slika 4.2.1. Područja istraživanja programu QGIS.

### 4.3. R programski jezik

R je besplatni softverski program otvorenog koda korišten za statističku analizu. RStudio je integrirano razvojno okruženje za R, programski jezik sastavljen od softverske aplikacije IDE (Integrated development environment) za statističko računanje i grafiku. Mogu se uvoziti datoteke bez potrebe pisanja koda. Podržava istraživanje i pismeno programiranje miješanjem koda i tekstualne dokumentacije. Izvorno podržava sučelje s Rmarkdownom, tako da se rezultati mogu prikazati u HTML, PDF ili Word formatu. RStudio podržava dijeljenje projekta povezivanjem s GitHubom i drugim sustavima za izradu verzija (Giorgi et al., 2022). Postoji više od 10000 R paketa koji mogu biti instalirani s naredbom *install.packages* („Naziv paketa“). Slika 4.3.1. prikazuje sučelje koje se dijeli na 4 prozora. Naredbe se pišu u prozorima R skripta (Rscript) ili u konzoli (Console), ali u R skripti (gore lijevo) mogu se spremirati, koristiti više puta ili čak korigirati. Nakon što se upiše naredba u konzolom prozoru (dolje lijevo), ona se izvrši i dobije se izlaz. Prozor (gore desno) koji sadrži karticu okruženja (Environment) obuhvaća sve varijable koje se stvaraju u R-u. U kartici povijesti (History) nalaze se sve prijašnje napisane naredbe. Četvrti prozor (dolje desno) ima karticu parcele (Plot) gdje se iscrtavaju grafovi i dijagrami. Do nje je kartica paketi (Packages) gdje su vidljivi svi dostupni paketi, a kartica za pomoć (Help) definira funkciju i provjerava njeno korištenje (R Learning Infrastructure Home Web Page, 2023., Patwal M., 2020., Kent State University, 2021., RStudio Wikipedia, 2023.).



Slika 4.3.1. RStudio.

### 4.3.1 R paketi

R paketi su zbirka funkcija, koje sadrže kodove, podatke, dokumentaciju u programskom jeziku R. Pomoću funkcije *install.packages()*, upisom željenog paketa u zagradu pokreće se instalacija zadanog paketa za daljnji rad. Nakon instalacije R paketi su spremljeni u direktorij pod nazivom *library()*. Pokretanjem naredbe *library()* učitavaju se paketi kako bi se funkcije, kodovi i podaci mogli upotrebljavati u R Studio-u. Tablica 4.3.1.1. opisuje R pakete korištene u istraživanju (Tutorialspoint, 2023., R package Wikipedia, 2023., Posit Support, 2023., Javatpoint-a, 2021., Spatial Data Science, 2023., CRAN-f, 2023., GitHub, 2023., CRAN-e, 2023., Rdr, 2023., CRAN-d, 2023., CRAN-c, 2023., CRAN-b, 2023., CRAN-a, 2023.).

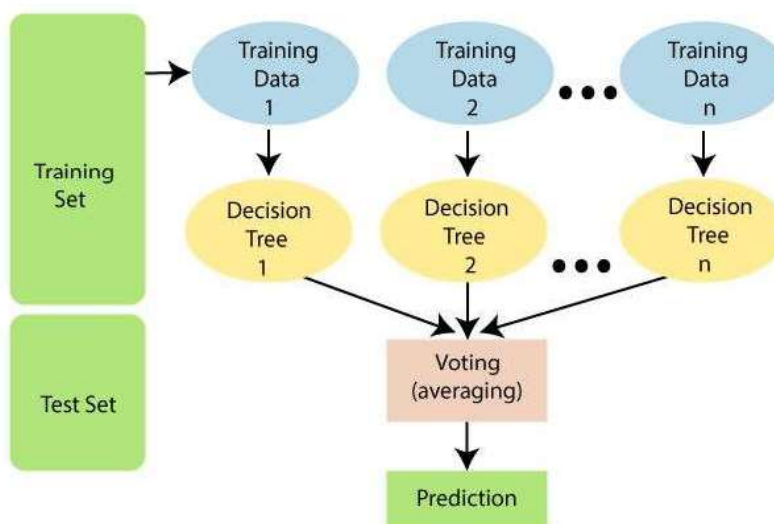


Tablica 4.3.1.1. Tablica paketa u R-u.

Naziv R paketa	Svrha paketa
<b>terra</b>	Paket se koristi za stvaranje, čitanje, manipuliranje i pisanje rasterskih podataka korištenih za razvoj specifičnih funkcija.
<b>sf</b>	Paket kodiranjem pojednostavljuje prikaz prostornog vektorskog podatka.
<b>dplyr</b>	Paket stvoren za brzu manipulaciju podataka gdje služi kao prečac za podskupove, sažimanje, preuređivanje i spajanje skupova podataka.
<b>ggplot2</b>	Paket služi za kreiranje kvalitetnih, elegantnih grafikona i slojevitih crteža.
<b>data.table</b>	Paket koji pri velikim podacima koristi alternativni način organiziranja skupova za brze operacije.
<b>caret</b>	Paket omogućuje obavljanje klasifikacije podataka.
<b>randomForest</b>	Paket za modeliranje podataka iz strojnog učenja.
<b>MLmetrics</b>	Paket procjene gubitka, rezultata i korisnosti koji mjeri regresiju, klasifikaciju i rangiranje podataka.
<b>plotly</b>	Paket pruža kvalitetne grafikone i online interakciju.
<b>LiblineaR</b>	Paket koji rješava velike linearne klasifikacije. Sadrži više klasnu klasifikaciju, unakrsnu provjeru valjanosti pri odabiru modela i brzu procjenu vjerojatnosti modela .
<b>grDevices</b>	Paket koji pruža podršku osnovnoj i mrežnoj grafici.
<b>RColorBrewer</b>	Paket nudi različite harmonije boja za karte i druge grafike.
<b>NeutralNetTools</b>	Paket s alatima vizualizacije i analize koji pomažu u interpretaciji modela s funkcijama iscertavanja, kvantificiranja važnosti varijable, provođenja analiza osjetljivosti i dobivanje jednostavnog popisa modela.
<b>mapview</b>	Paket koji brzo izrađuje prostorne karte sa ili bez pozadinskih karata, obuhvaća metode za vizualizaciju rasterskih slika u pravim i lažnim bojama i graničnih okvira.
<b>leaflet</b>	Paket kreira višestruke web-karte s brošurama koje imaju sinkroniziranu kontrolu pomicanja, povećavanja, usporedbe atributa s različitom geometrijom i mogućnost sinkroniziranja bilo koje kombinacije karata.

#### 4.4. Metode strojnog učenja za klasifikaciju

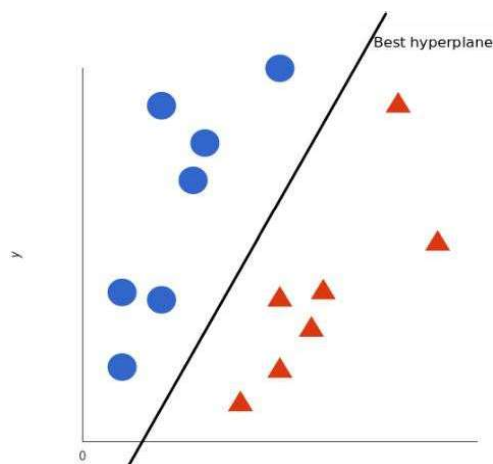
**Random Forest** predstavlja nadzirani algoritam, koji se koristi za probleme klasifikacije i regresije. Nudi mogućnost kombiniranja više klasa za rješavanje kompleksnih problema i unapređivanje izvedbe modela. Metoda koja koristi više stabala odlučivanja za donošenje predviđanja te donosi konačni rezultat. Karakteristike *randomForest* algoritma su veća točnost, važnost procjena klasifikacije, pouzdana obrada podataka koji nedostaju, kratkotrajnost obuke te korištenje za slijedeću primjenu s ostalim podacima. R nudi praktičan i lako iskoristiv algoritam. Za korištenje *randomForest* u R-u mora se učitati skup podataka, zatim se skup podataka podijeli za treniranje i testiranje. Nakon treniranja *randomForest*, dobivaju se predviđanja na skupu za testiranje (Slika 4.4.1.). Na kraju se procjenjuje izvedba modela upotrebom raznovrsnih mjera kao što su preciznost, točnost (Simplilearn, 2023., Javatpoint-b, 2021.).



Slika 4.4.1. Prikaz rada RF-a (Javatpoint-b, 2021.).

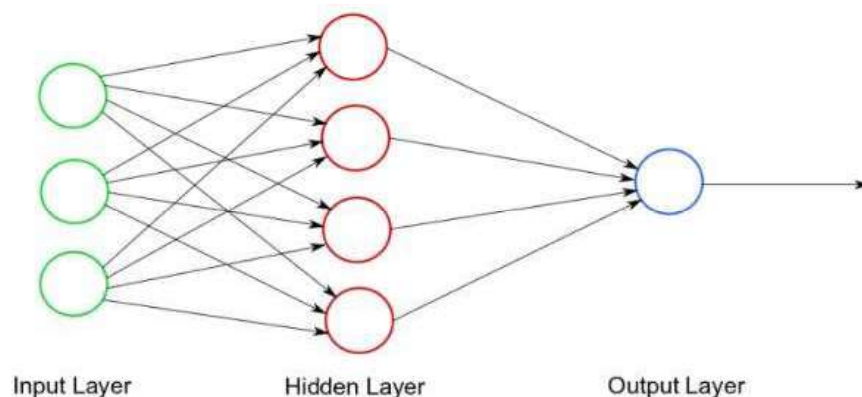
**Support vector machine** predstavlja algoritam s potpunim vektorima za probleme klasifikacijske i regresijske. Pruža pronalaženje hiper-ravnine koja najbolje razdvaja podatke u prostoru. Ima svrhu promatranja i izrade granica danih grupa podataka za predviđanje nadolazećih grupa koje se temelje na njihovim mjerenjima. Dobivaju se “klase“ razdvajanjem različitih skupina (Slika 4.4.2.). Prednosti *SVM-a* su velika dimenzionalnost, oponašanje bilo kojeg oblika i prilagodba za svaku zamišljenu klasifikaciju. U R Studio-u nudi praktičan i lako iskoristiv algoritam. Za korištenje *SVM* u R-u mora se učitati skup podataka, zatim se skup podataka podijeli za treniranje i testiranje. Nakon treniranja *SVM* modela, dobivaju se predviđanja na skupu za testiranje. Na kraju se procjenjuje izvedba modela

upotrebom raznovrsnih mjera kao što su preciznost, točnost (UC Bussines Analytics R Programming Guide, 2023., Datacamp, 2018., Geeksforgeeks, 2021.).



Slika 4.4.2. Prikaz rada SVM-a (Datacamp, 2018.).

**Neural Network** inspirirana je radom neurona u ljudskom mozgu za rješavanje raznih problema. Imitira međusobnu komunikaciju neurona za donošenje odluke. Slika 4.4.3. prikazuje mrežu sastavljenu od ulaznog (neobrađeni podaci), skrivenog (obrada podataka) i izlaznog sloja (dobivanje rezultata). Provodi se obuka neuronskih mreža za klasifikaciju podataka. Za uspješne rezultate klasifikacije modela potreban je niz ponavljanja. U R programskom jeziku moguća je izvedba neuronske mreže upotrebom “*nnet*“ paketa, te analizu rezultata i vizualizaciju. Prednosti neuronske mreže su praktičnost za rješavanje problema regresije i klasifikacije, modeliranje nelinearnih podataka s ogromnim brojem ulaza, velika brzina predviđanja, treniranje različitim brojevima ulaza i slojeva te najbolje funkcioniraju s većom količinom podataka (Datacamp, 2023., Geeksforgeeks, 2020., Packt, 2023.).



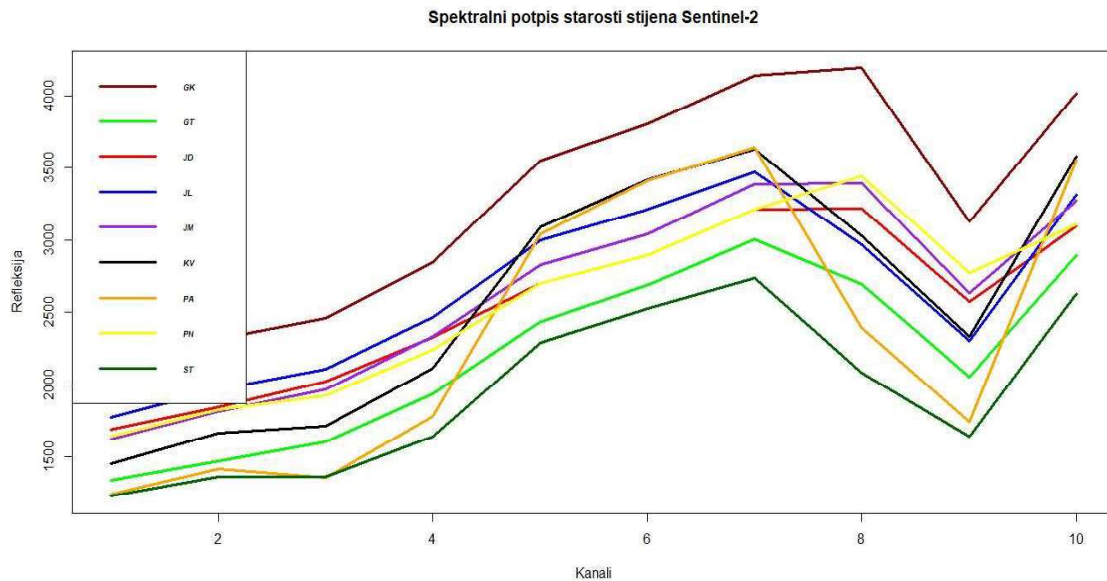
Slika 4.4.1. Prikaz rada Neural Network (Geeksforgeeks, 2020.).

## 5. Analiza spektralnog potpisa satelitske misije Sentinel-2

Grafički prikaz spektralnog potpisa za multispektralnu snimku Sentinel-2 (15. 04. 2022.) prikazuje refleksiju svake klase za pojedini kanal. Izrađene su dvije detaljne analize spektralnog potpisa pomoću satelitske misije Sentinel-2. Prva spektralna analiza usmjerena je na određivanje starosti, a druga na identifikaciju različitih vrsta stijena u promatranom području. U obje analize, koristili su se podaci dobiveni kroz deset različitih kanala. Svaki kanal predstavlja određeni dio elektromagnetskog spektra i nosi informacije o različitim aspektima promatranog područja. Tablice s prikupljenim podacima su učitane i pretvorene u matrice koje sadrže odgovarajuće klase. Ove matrice su ključne za stvaranje grafičkih prikaza krivulja, čime se olakšava interpretacija spektralnih potpisa. Klase su definirane različitom bojom kako bi se jasno vizualizirali rezultati analize. Korišteni su isti kodovi za dobivanje spektralnih potpisa Sentinel-2 za starost i vrstu stijena. U nastavku slijedi kod:

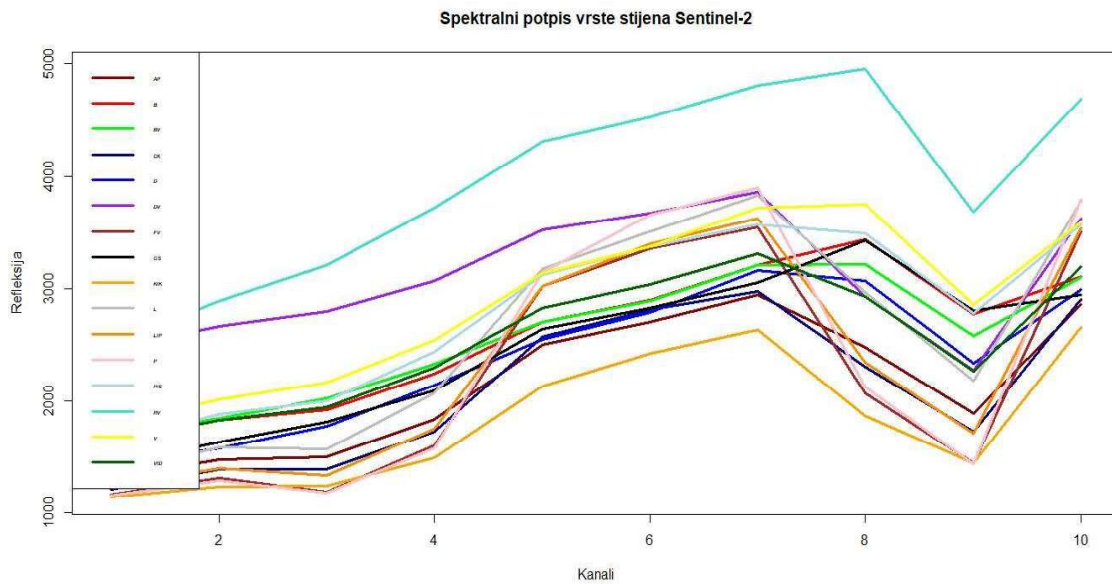
```
set.seed(555)
p <- st_sample( samp1, 1000, method= "regular")
p_sf <- st_sf(p)
p_at <- st_join(p_sf, samp1)
#df <- terra::extract(as(sentinel, "Raster"), p_at)
df <- terra::extract(as(s1, "Raster"), p_at)
df <- as.data.frame(df)
df$starostID <- p_at$starostID
df$starost <- p_at$starost
ms <- aggregate(df, list(p_at$starost), mean, na.rm=TRUE)
#mjenjanje brojeva
ms <- ms[-12]
rownames(ms) <- ms[,1]
ms <- ms[,-1]
mycolor <- c('darkred', 'green', 'red', 'blue', 'purple',
'black', 'orange', 'purple', 'drakgreen')
ms <- as.matrix(ms)
plot(0, ylim=c(min(ms),max(ms)), xlim = c(1,10), type='n',
xlab="Kanali", ylab = "Refleksija")
for (i in 1:nrow(ms)){lines(ms[i,], type = "l", lwd = 3, lty =
1, col = mycolor[i])}
title(main="Spektralni potpis starosti stijena Sentinel-2",
font.main = 2)
legend("topleft", rownames(ms), cex=0.55, text.font=4,
col=mycolor, lty = 1, lwd =3, bty = "y")
```

Dobiven je grafički prikaz spektralnog potpisa starosti stijena za multispektralne snimke Sentinel-2 (Slika 5.1.). Sve krivulje imaju karakterističan oblik koji odgovara opisu klase. Krivulje devet klasa definirane su različitim bojama, srednji trijas (tamno zeleno), gornji trijas (zeleno), jura lijas (plavo), jura doger (tamno crveno), jura malm (ljubičasto), gornja kreda (smeđe), paleogen (narančasto), paleogen-neogen (žuto), kvartar (crno). Svaka krivulja ima svoj karakterističan oblik, postižu različite trendove u pojedinim kanalima. Najizraženija je krivulja gornje krede (smeđe) zbog povišenija u odnosu na većinu krivulja koje su slične. Povećavanje trendova u ovom slučaju započinje od četvrtog kanala, a nagli pad uočen je za deveti kanal.



Slika 5.1. Spektralni potpis starosti stijena Sentinel-2.

Spektralni potpis vrste stijena za multispektralne snimke Sentinel-2 prikazan je za deset kanala (Slika 5.2.). Krivulje šesnaest klasa definirane su različitim bojama, amfibolski porfirit (smeđe), vapnenci (žuto), vapnenci i dolomiti (tamno zeleno), dolomiti (plavo), crveni klasititi (tamno plavo), konglomerati i klastiti (narančasto), dolomitizirani vapnenci (ljubičasto), brečoliki vapnenci (zeleno), rudistni vapnenci (tirkizno), foraminiferski vapnenci (tamno crveno), lapori (sivo), lapori i pješčenjaci (tamno narančasto), breče (crveno), glaciofluvijalni sedimenti (crno), pijesak i šljunak (svijetlo plavo), pijesci (rozo). Sve krivulje imaju karakterističan oblik koji odgovara opisu klase. Krivulje koje se ističu su rudistni vapnenci (tirkizno) i dolomitizirani vapnenci (ljubičasto) zbog povišenija u odnosu na većinu krivulja koje su slične. Povećavanje trendova u ovom slučaju započinje od četvrtog kanala, a nagli pad uočen je za deveti kanal.



Slika 5.2. Spektralni potpis vrste stijena Sentinel-2.

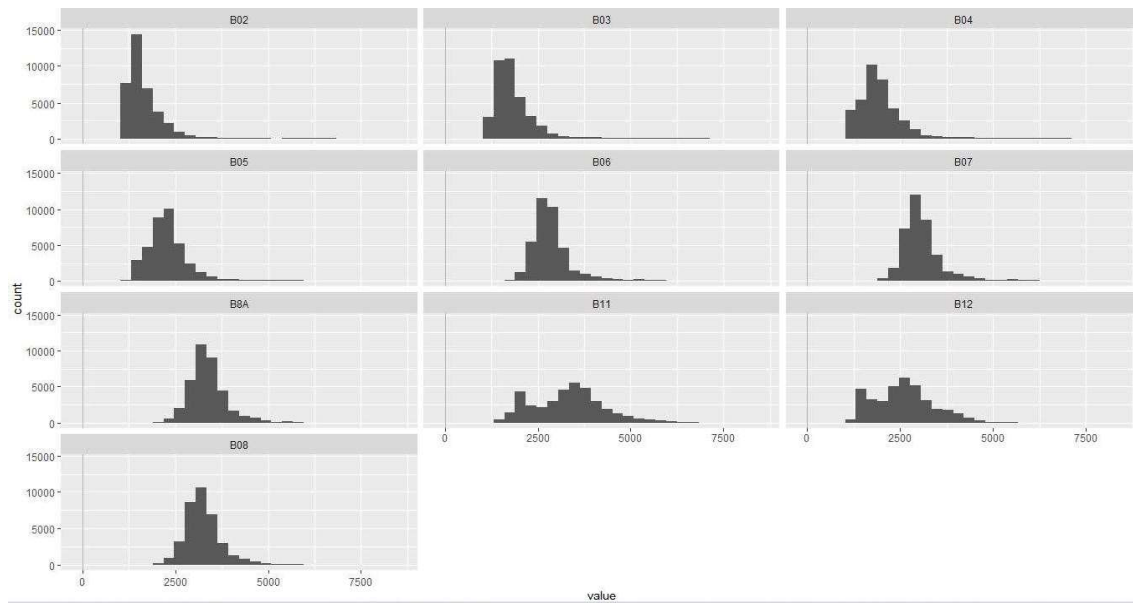
Za potrebe istraživanja izabran je Sentinel-2, kao reprezentativni model zbog veće prostorne rezolucije i pouzdanosti. Pošto se istraživanja obavljaju na prostorima Republike Hrvatske, koja je članice Europske Unije, pogodnosti koje nudi europska svemirska agencija Copernicus, satelita Sentinel-2, su dobivanje čistih i preciznijih snimaka za potrebe istraživanja klasifikacije litologije.

## 6. Obrada podatka metodama strojnog učenja .

Primjenom Copernicus Open Access Hub-a preuzimaju se multispektralne snimke satelita Sentinel-2, za odabrano područje istraživanja. Definiraju se imena 10 kanala satelitskih snimaka. Koristeći GADM, sustav koji nudi karte i prostorne podatke svih zemalja, učitavaju se granice općina (level 2) u Hrvatskoj iz baze podataka. Izdvajaju se područja istraživanja (općina Senj, otok Rab, otok Sveti Grgur i Goli otok) iz većeg skupa geoprostornih podataka. Dobiva se izdvojeno područje na temelju zadanih granica koje se koriste za maskiranje vrijednosti *Raster* objekta izvan područja definiranog drugim *Raster* objektom. Rezultat je multispektralni rasterski objekt koji prikazuje područje istraživanja. Preklopljen je sa vektorskim objektom koji predstavlja poligone (starosti i vrste stijena) klasa. Korištenjem funkcija za izdvajanjem pojedinačnih informacija o vrijednosti piksela stvorena je tablica koja sadrži za svaki piksel sve vrijednosti kanala i pridodanu vrijednost pripadajuće klase. Zbog velike količine brojčano dobivenih podataka za poligone starosti i vrste stijena, prikazani su samo rezultati za poligon starost stijena. Prikaz par elemenata i struktura dobivene tablice prikazano je u tablici 6.1. u kojoj se nalazi ukupno 36886 elemenata. Ta tablica je ulazni podatak za korištene algoritme strojnog učenja u ovom radu. Na slici 6.1. prikazani su histogrami za svaki kanal multispektralnog snimka.

Tablica 6.1. Prikaz vrijednosti dobivene tablice.

	<b>B02</b>	<b>B03</b>	<b>B04</b>	<b>B05</b>	<b>B06</b>	<b>B07</b>	<b>B8A</b>	<b>B11</b>	<b>B12</b>	<b>B108</b>	<b>Klasa</b>
1:	1164	1320	1273	1580	2583	2781	3048	1957	1491	3190	ST
2:	1168	1330	1272	1610	2604	2915	3088	1925	1476	3098	GT
3:	1158	1294	1244	1555	2483	2705	2946	1879	1466	2730	JL
4:	1158	1292	1258	1575	2452	2724	2945	1923	1483	2840	JD
5:	1219	1397	1346	1746	2591	2944	3168	2271	1673	3238	JM
-----											
36882:	1804	2057	2286	2608	3065	3285	3638	3965	3104	3470	GK
36883:	1726	1995	2191	2628	3205	3332	3552	3931	3058	3666	PA
36884:	1559	1816	2010	2606	3181	3330	3622	3901	3031	3422	PN
36885:	1526	1742	1857	2191	2845	3131	3347	3491	2646	3360	KV
36886:	1623	1930	2137	2548	3239	3454	3755	3840	2950	3792	ST



Slika 6.1. Prikaz histograma multispektralnih snimaka.

Korištenjem funkcije “*createDataPartition()*” iz paketa *caret*, radi se podjela početnog skupa podataka na skupu za treniranje i skupu za testiranje. U skupu za treniranje “*dt\_train*” samo odabrani redci izvornog skupa podataka se koriste za treniranje modela (Tablica 6.2.), a u skupu za testiranje “*dt\_test*” se samo preostali redci izvornog skupa podataka koriste za testiranje modela (Tablica 6.3.). U nastavku slijedi kod:

```
set.seed(321)
idx_train <- createDataPartition(dt$class,
                                p = 0.7,
                                list = FALSE)

dt_train <- dt[idx_train]
dt_test <- dt[-idx_train]

table(dt_train$class)
table(dt_test$class)
```

Tablica 6.2. Količina treninga podataka klasa.

1. ST	2. GT	3. JL	4. JD	5. JM	6. GK	7. PA	8. PN	9. KV
764	957	2703	4023	5004	3479	977	7058	859



Tablica 6.3. Količina testa podataka klasa.

1. ST	2. GT	3. JL	4. JD	5. JM	6. GK	7. PA	8. PN	9. KV
327	409	1158	1723	2144	1491	418	3024	368

Kreiranju se preklapni skupovi za slučajni generator brojeva koji se kasnije koriste u postupku treniranja i potvrde modela. Pomoću naredbe “*n\_folds*” stvara se broj preklapnih skupova i generira listu za svaki od njih. Funkcija “*trainControl()*” omogućuje kontroliranje procesa učenja i procjene modela. Koristi se metoda za određivanje procjene pouzdanost i izvedbe modela (strategija unakrsne potvrde) “*cv*”, broja presavijanja “*number*”, načina pretrage hiperparametara “*search*” i spremanja predikcija “*savePredictions*”. Za izračunavanje vjerojatnosti klasa za procjenu izvedbe modela koristi se “*classProbs*”, a “*seeds*” za postupak ponavljanja nasumičnosti procesa. U nastavku slijedi kod:

```
n_folds <- 10
set.seed(321)
folds <- createFolds(1:nrow(dt_train), k = n_folds)
seeds <- vector(mode = "list", length = n_folds )
for(i in 1:n_folds) seeds[[i]] <- sample.int(1000, n_folds)
seeds[n_folds + 1] <- sample.int(1000, 1)
ctrl <- trainControl(summaryFunction = multiClassSummary,
                      method = "cv",
                      number = n_folds,
                      search = "grid",
                      classProbs = TRUE,
                      savePredictions = TRUE,
                      index = folds,
                      seeds = seeds)
```

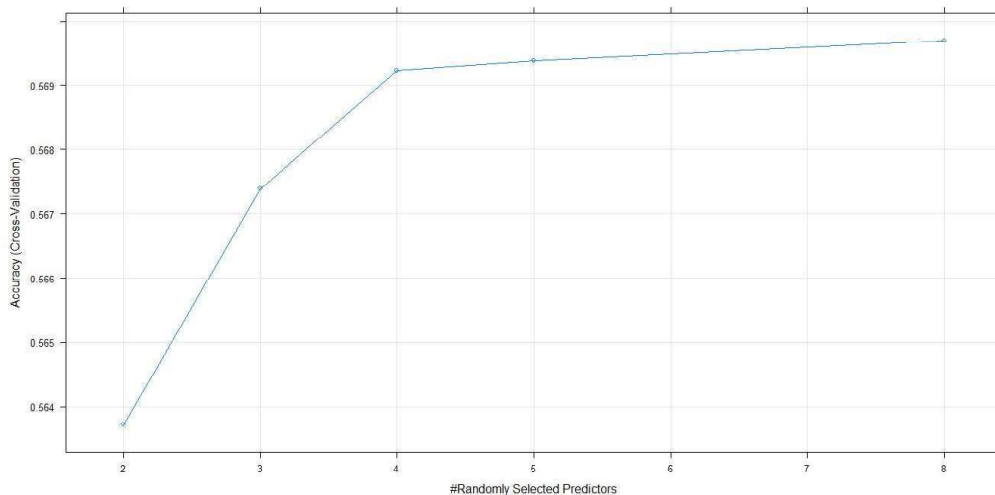
### 6.1. Korištenje metode *randomForest* (RF) iz paketa CARET

Primjenjuje se metoda *randomForest* (RF). Varijabla “*model\_rf*” koristi funkciju „*train()*” za treniranje modela slučajne šume, obavlja optimizaciju hiperparametara modela pomoću unakrsne provjere i računa važnost prediktorskih varijabli. Nakon treniranja, vraća model i njegovu procjenu izvedbe. Funkcija “*model\_rf\$times\$everything*” odnosi se na vrijeme izračuna potrebno za svaku iteraciju procesa preklapanja, uključujući prilagodbu i predviđanje modela. Rezultat je grafički prikaz koji predočava rast točnosti s povećanjem nasumične predikcije (Slika 6.2.). U nastavku slijedi kod:

```

model_rf <- caret::train(class ~ . , method = "rf", data =
dt_train,
                        importance = TRUE,
                        tuneGrid = data.frame(mtry = c(2, 3, 4,
5, 8)),
                        trControl = ctrl)
saveRDS(model_rf, file = "~/model_rf3.rds")
model_rf
model_rf$times$everything
plot(model_rf)

```



Slika 6.2. Grafički prikaz predikcije RF-a.

Funkcija “*confusionMatrix()*” za varijablu *cm\_rf*, na temelju modela slučajne šume stvara matricu zabune, koja prikazuje broj točnih i netočnih predikcija za svaku klasu ciljne varijable (Tablica 6.1.1.). Objekt *model\_rf\$finalModel* predstavlja naučeni model slučajne šume, koji se koristi za predviđanje ciljne varijable za nove podatke pomoću naredbe “*predict()*”. U nastavku slijedi kod:

```

cm_rf <- confusionMatrix(data = predict(model_rf, newdata =
dt_test), as.factor(dt_test$class))
cm_rf
model_rf$finalModel

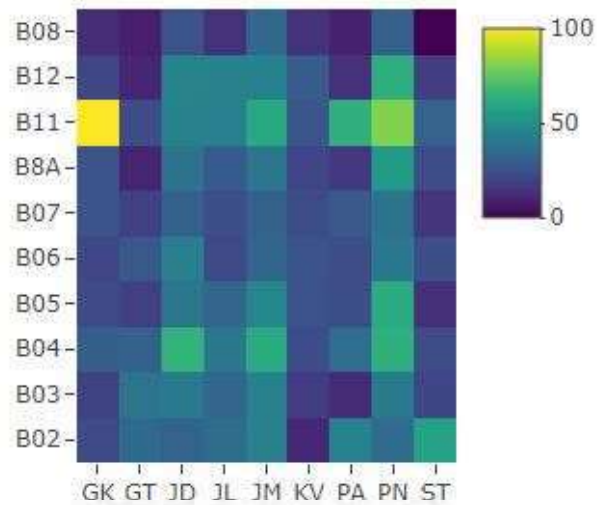
```

Tablica 6.1.1. Statistički prikaz *random forest* klasifikacije.

<b>Statistics by Class:</b>	<b>Class: ST</b>	<b>Class: GT</b>	<b>Class: JL</b>	<b>Class: JD</b>	<b>Class: JM</b>	<b>Class: GK</b>	<b>Class: PA</b>	<b>Class: PN</b>	<b>Class: KV</b>
Sensitivity	0.51682	0.27873	0.49655	0.62914	0.5956	0.8578	0.73206	0.8251	0.51359
Specificity	0.99320	0.99211	0.96153	0.92665	0.9026	0.9844	0.99145	0.8544	0.99327
Pos Pred Value	0.69835	0.57576	0.60146	0.61278	0.5951	0.8957	0.77078	0.6808	0.72414
Neg Pred Value	0.98540	0.97285	0.94231	0.93124	0.9028	0.9780	0.98950	0.9285	0.98343
Prevalence	0.02956	0.03697	0.10468	0.15576	0.1938	0.1348	0.03779	0.2734	0.03327
Detection Rate	0.01528	0.01031	0.05198	0.09799	0.1154	0.1156	0.02766	0.2255	0.01709
Detection Prevalence	0.02188	0.01790	0.08642	0.15992	0.1940	0.1291	0.03589	0.3313	0.02359
Balanced Accuracy	0.75501	0.63542	0.72904	0.77789	0.7491	0.9211	0.86175	0.8398	0.75343

Funkcija `varImp()` pretvara izračunate vrijednosti u matricu. Dobiva se grafikon *heatmap* koji prikazuje značajnost svake komponente za klasifikaciju ciljne varijable (Slika 6.3.). Tamnija boja (plava) označava bolju klasifikaciju za određeni kanal, dok svjetlija boja (žuta) označava lošiju. U nastavku slijedi kod:

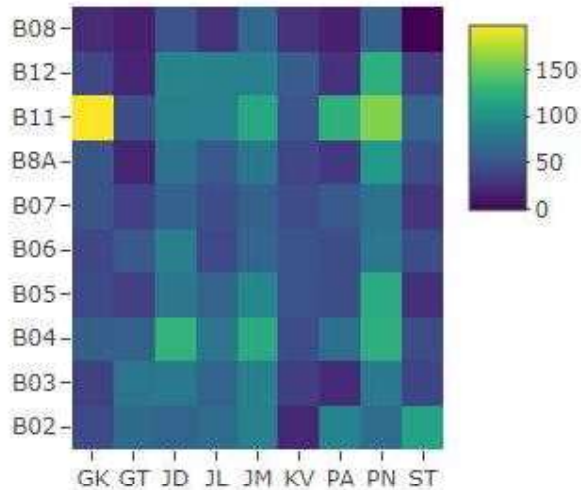
```
caret::varImp(model_rf)$importance %>%
  as.matrix %>%
  plot_ly(x = colnames(.), y = rownames(.), z = ., type =
"heatmap",
  width = 350, height = 300)
```



Slika 6.3. Prikaz *heatmap1*.

Izračunavanje važnost značajki u modelu slučajne šume izvodi se funkcijom “*importance()*” iz paketa *randomForest*. Vrijednosti se pretvaraju u matricu i stvara se *heatmap* grafikom naredbom “*plot\_ly()*” iz paketa *plotly* (Slika 6.4.). U nastavku slijedi kod:

```
randomForest::importance(model_rf$finalModel) %>%
  .[, - which(colnames(.) %in% c("MeanDecreaseAccuracy",
    "MeanDecreaseGini"))] %>%
  plot_ly(x = colnames(.), y = rownames(.), z = ., type =
    "heatmap",
    width = 350, height = 300)
```



Slika 6.4. Prikaz *heatmap*2.

## 6.2. Korištenje metode *Support Vector Machine* (SVM) iz paketa CARET

Sljedeća metoda koja se primjenjuje, stvara mrežu parametara za algoritam koji se zove *Support Vector Machine* (SVM). Koristi se za pronalaženje najbolje vrijednosti parametara, tako da nudi bolje rezultate u analizi podataka. Naredba “*expand.grid()*” kombinira tri vrijednosti za parametar “*cost*” (0.2, 0.5, 1) s dvije vrijednosti za parametar “*Loss*” (“L1” i “L2”) za stvaranje mreže sa šest kombinacija. Obučavanje modela se izvodi pomoću skupa podataka i linearnog SVM algoritma. Cilj je pronaći najbolje vrijednosti parametara za algoritam kako bi se model bolje prilagodio podacima. Za podešavanje hiperparametara se koristi mreža “*svm\_grid*”, za izvedbu modela upotrebljava se kontrola treniranja *trControl*, a za obuku modela koristi se funkcija “*caret::train()*”. Podskup podataka sadrži informacije o vremenu izvođenja svakog koraka u procesu obuke SVM modela. Naredba “*plot(model\_svm)*” predstavlja grafički prikaz izvedbe modela koji je obučen pomoću “*model\_svm*” objekta (Slika 6.2.1.). Izračunavaju se matrica zabune koje prikazuje koliko je model predvidio ispravno i pogrešno za svaku klasu, na temelju predviđanja modela na testnom skupu podataka i stvarnih vrijednosti ciljne varijable “*dt\_test\$class*” (Tablica 6.2.1.). U nastavku slijedi kod:

```
svm_grid <- expand.grid(cost = c(0.2, 0.5, 1),
                      Loss = c("L1", "L2"))
model_svm <- caret::train(class ~ . , method = "svmLinear3",
                          data = dt_train,
                          allowParallel = TRUE,
                          tuneGrid = svm_grid,
                          trControl = ctrl)

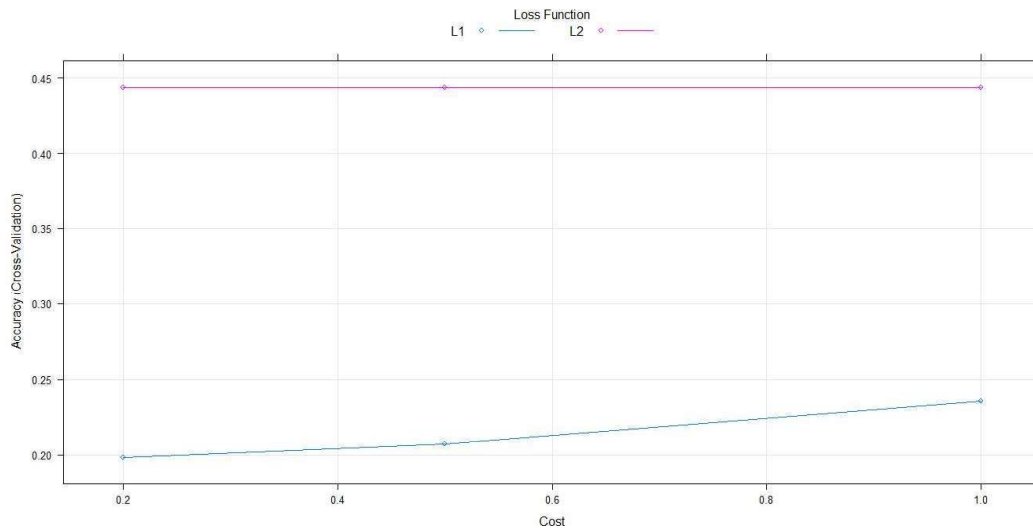
saveRDS(model_svm, file = "~/model_svm3.rds")
```

```

model_svm$times$everything
plot(model_svm)

cm_svm <- confusionMatrix(data = predict(model_svm, newdata =
dt_test), as.factor(dt_test$class))
cm_svm

```



Slika 6.2.1. Grafički prikaz modela SVM-a.

Tablica 6.2.1. Statistički prikaz *svm* klasifikacije.

Statistics by Class:	Class: ST	Class: GT	Class: JL	Class: JD	Class: JM	Class: GK	Class: PA	Class: PN	Class: KV
Sensitivity	0.00000	0.00000	0.10967	0.19269	0.40159	0.68880	0.45694	0.7778	0.06522
Specificity	1.00000	1.00000	0.97345	0.94496	0.81487	0.92613	0.97943	0.6555	0.99766
Pos Pred Value	-	-	0.32564	0.39243	0.34275	0.59227	0.46585	0.4593	0.48980
Neg Pred Value	0.97044	0.96303	0.90339	0.86384	0.84994	0.95026	0.97869	0.8869	0.96876
Prevalence	0.02956	0.03697	0.10468	0.15576	0.19382	0.13479	0.03779	0.2734	0.03327
Detection Rate	0.00000	0.00000	0.01148	0.03001	0.07783	0.09284	0.01727	0.2126	0.00217
Detection Prevalence	0.00000	0.00000	0.03526	0.07648	0.22708	0.15675	0.03706	0.4629	0.00443
Balanced Accuracy	0.50000	0.50000	0.54156	0.56882	0.60832	0.80747	0.71818	0.7166	0.53144

### 6.3. Korištenje metode *Neural Network* (NNET) iz paketa CARET

Posljednja metoda koja se primjenjuje je *Neural Network* (NNET). Dobivaju se informacije o treniranom modelu za procjenu njegove izvedbe. Stvara se mreža svih kombinacija vrijednosti parametara na temelju definiranih argumenta. Kombiniraju se tri veličine *size* koje predstavljaju broj neurona u sloju (5, 10 ili 15) s tri stope degradacije *decay* koje određuju brzinu da se težine mreže smanjuju tijekom treniranja (0.001, 0.01 ili 0.1). Model se trenira na podacima za sve dostupne značajke “*class*“. Varijablom *nnet\_grid* postavljeni su parametri modela koje će se ispitati. Izvodi se unakrsna validacija tijekom treniranja i procjene izvedbe modela. Naredbom “*plot(model\_nnet)*“ dobivaju se različiti dijagrami koji prikazuju izvedbu modela (Slika 6.3.1.). Zatim se ponovno izračunavaju matrice zabune predviđanjem modela i stvarnih vrijednosti iz testnog skupa (Tablica 6.3.1.). U nastavku slijedi kod:

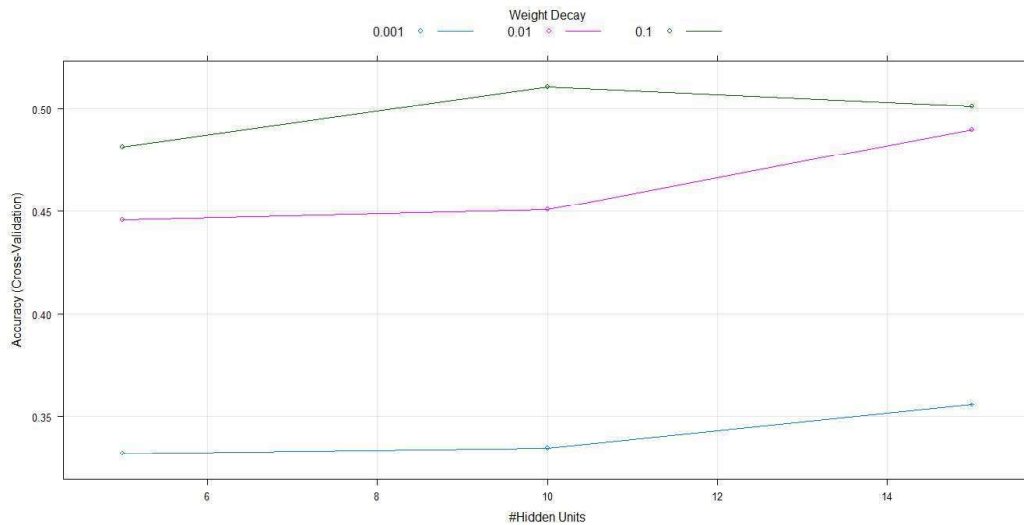
```
nnet_grid <- expand.grid(size = c(5, 10, 15),
                        decay = c(0.001, 0.01, 0.1))

model_nnet <- train(class ~ ., method = 'nnet', data = dt_train,
                    importance = TRUE,
                    maxit = 1000,
                    allowParallel = TRUE,
                    tuneGrid = nnet_grid,
                    trControl = ctrl)

saveRDS(model_nnet, file = "~/model_nnet3.rds")

model_nnet$times$everything
plot(model_nnet)

cm_nnet <- confusionMatrix(data = predict(model_nnet,
newdata = dt_test), as.factor(dt_test$class))
cm_nnet
```



Slika 6.3.1. Grafički prikaz modela *Neural Network*.

Tablica 6.3.1. Statistički prikaz *nnet* klasifikacije.

Statistics by Class:	Class: ST	Class: GT	Class: JL	Class: JD	Class: JM	Class: GK	Class: PA	Class: PN	Class: KV
Sensitivity	0.53517	0.031785	0.29879	0.48230	0.43097	0.8739	0.59809	0.7629	0.37228
Specificity	0.98510	0.997184	0.94134	0.90374	0.86488	0.9763	0.98403	0.8268	0.98953
Pos Pred Value	0.52239	0.302326	0.37325	0.48035	0.43401	0.8516	0.59524	0.6237	0.55020
Neg Pred Value	0.98583	0.964062	0.91988	0.90441	0.86343	0.9803	0.98421	0.9026	0.97864
Prevalence	0.02956	0.036973	0.10468	0.15576	0.19382	0.1348	0.03779	0.2734	0.03327
Detection Rate	0.01582	0.001175	0.03128	0.07512	0.08353	0.1178	0.02260	0.2086	0.01238
Detection Prevalence	0.03028	0.003887	0.08380	0.15639	0.19246	0.1383	0.03797	0.3344	0.02251
Balanced Accuracy	0.76013	0.514484	0.62006	0.69302	0.64793	0.9251	0.79106	0.7949	0.68090

Funkcija za interpolaciju boja `grDevices::colorRampPalette`, upotrebljava naredbu `“brewer.pal”` iz paketa `RColorBrewer` kako bi se kreirala paleta s odabranim imenom 10 različitih boja za potrebe vizualizacije podataka i dobivanja grafova. Računanje važnosti značajki u neuronskoj mreži izvodi se funkcijom `“garson(model_nnet)”`, treniranoj na skupu podataka, a `“scale_y_continuous”` i `“scale_fill_gradientn”` su naredbe za prilagođavanje osi y i gradijenta boja u izlaznom grafu. Uvježbani `model_nnet` omogućuje prikaz neuronske mreže (Slika 6.3.2.). Koriste se boje koje odgovaraju bojama

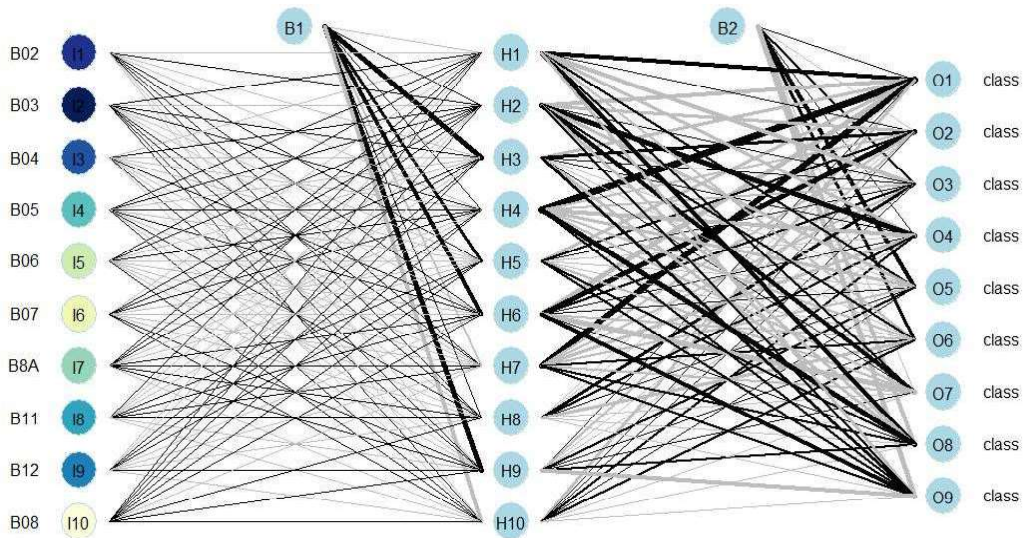


iz funkcije “cols“, poretka sortiranog prema važnosti neurona za dobivanje dijagrama neuronske mreže (Slika 6.3.3.). Za učeni model, kanali imaju pouzdan utjecaj na ciljani rezultat, koji se smanjuje s desna na lijevo. Poveznice između neuronske mreže i dijagrama su iste boje kanala, tamnije boje ukazuju na pouzdanije kanale u odnosu na svjetlije boje. U nastavku slijedi kod:

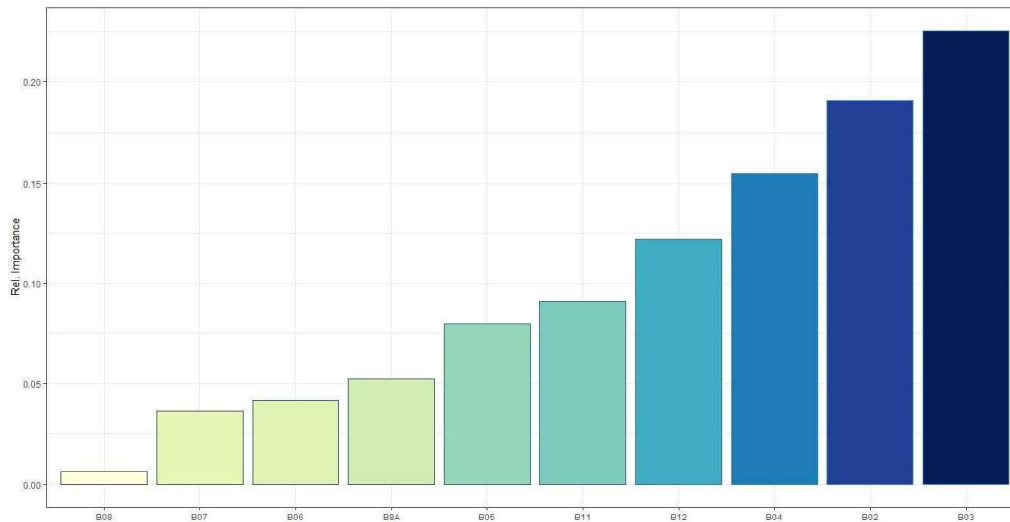
```
cols <- grDevices::colorRampPalette(colors = brewer.pal(n = 9,
name = "YlGnBu"))(10)

garson(model_nnet) +
  scale_y_continuous('Rel. Importance') +
  scale_fill_gradientn(colours = cols)

cols_rank_import <- cols[rank(garson(model_nnet, bar_plot =
FALSE)$rel_imp)]
plotnet(model_nnet, circle_col = list(cols_rank_import,
'lightblue'))
```



Slika 6.3.2. Prikaz neuronske mreže.



Slika 6.3.3. Dijagram neuronske mreže.

#### 6.4. Usporedba dobivenih modela

Stvara se lista modela za usporedbu izvedbe koja uključuje tri različita modela “rf” (Random Forest), “svm” (SVM), “nnet” (Neural Network). Zatim se funkcijom `caret::resamples` izvodi postupak unakrsne validacije za svaki model u listi (Slika 6.4.1.). Naredba “`bwplot()`” crta (*box-and-whisker*) grafove za svaki model u objektu `resamples`. Funkcija `resamples$metrics` za svaki model vraća tablicu koja sadrži različite vrijednosti. Računaju se razlike između svih kombinacija modela u listi funkcijom “`diff`” iz paketa `caretEnsemble`, a izračunavanje p-vrijednosti i konstruiranje tablice sažetka izvodi se naredbom “`summary`”. Rezultati su pouzdaniji što su bliži vrijednosti jedan za određene funkcije. Za recall i sensitivity funkciju najbliži je model `nnet`, dok model `svm` nedostaje u F1, Pos Pred Value, Precision. U nastavku slijedi kod:

```

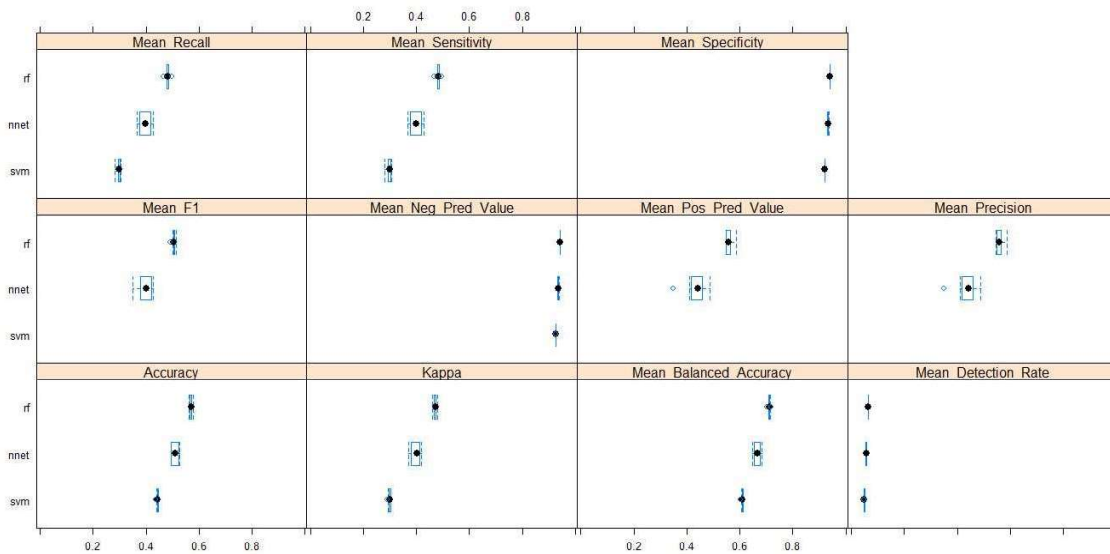
model_list <- list(rf = model_rf, svm = model_svm, nnet =
model_nnet)
resamples <- caret::resamples(model_list)

bwplot(resamples)

resamples$metrics

t_tests <- resamples %>%
  diff(metric = "Accuracy") %>%
  summary
t_tests

```



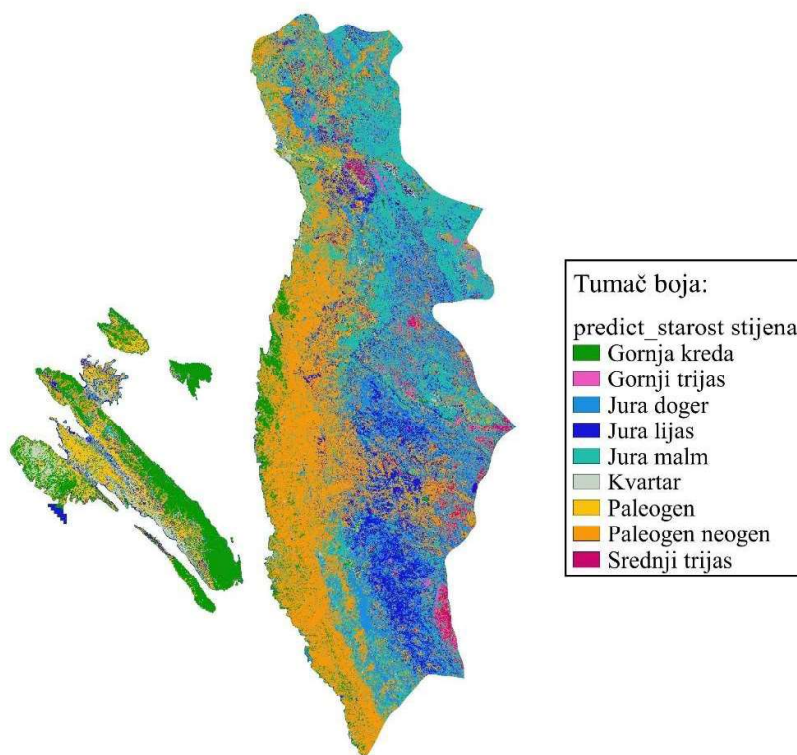
Slika 6.4.1. Prikaz rezultata unakrsne validacije.

Vrijeme izvršavanja tri modela koji su trenirani i pohranjeni kao *model\_rf*, *model\_svm* i *model\_nnet*, mjeri se i prikazuje na izlazu funkcijom “*system.time()*“. Primjena modela na piksele iz *sm* skupa podataka izvodi se naredbom *raster::predict*. Na kraju se dobiva grafički prikaz tri modela i njihovi tifovi koji se pohranjuju za daljnju primjenu i istraživanja. U nastavku slijedi kod:

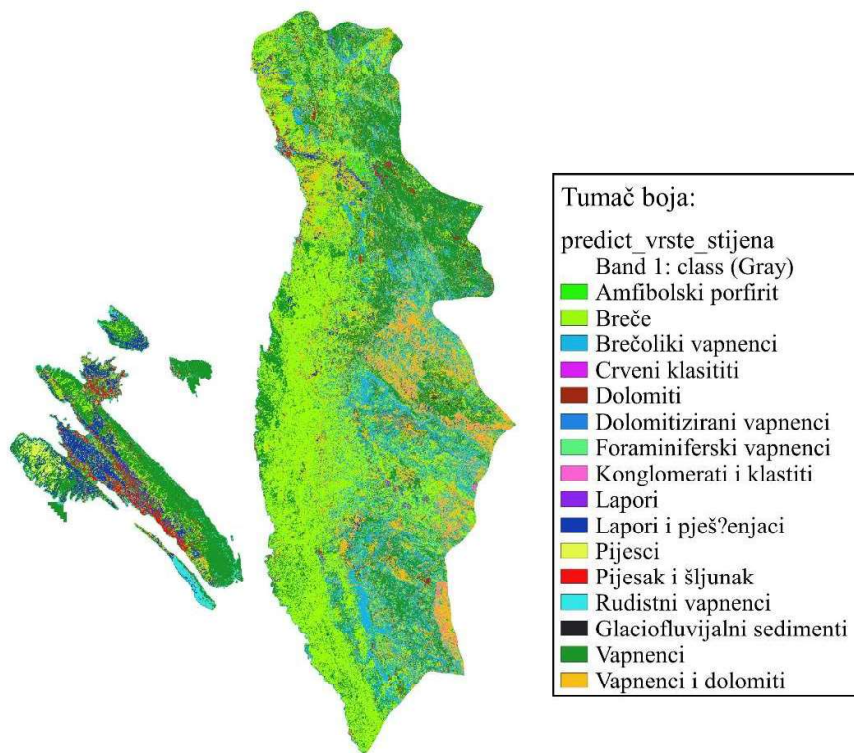
```
system.time({
  predict_rf <- raster::predict(object = sm,
                                model = model_rf, type = 'raw',
                                na.rm=TRUE)
  predict_svm <- raster::predict(object = sm,
                                 model = model_svm, type =
'raw', na.rm=TRUE)
  predict_nnet <- raster::predict(object = sm,
                                  model = model_nnet, type =
'raw', na.rm=TRUE)})
```

## 7. Rezultati i rasprava korištenjem Sentinel-2

Grafički prikazi tri modela *model\_rf*, *model\_svm* i *model\_nnet* prikazani su u QGIS-u za pregled dobivene klasifikacije za područje općine Senj, otoka Raba, otoka Sveti Grgur i Golog otoka te određivanje točnosti na temelju zadane starosti i vrste stijena. Za vremensko razdoblje stijena, klasificirano je devet grupa koje su definirane različitim bojama, srednji trijas (tamno rozo), gornji trijas (rozo), jura lijas (tamno plavo), jura doger (plavo), jura malm (tirkizno), gornja kreda (zeleno), paleogen (žuto), paleogen-neogen (narančasto), kvartar (sivo). Za vrstu stijena, klasificirano je šesnaest grupa koje su definirane različitim bojama, amfibolski porfirit (kričavo zeleno), vapnenci (tamno zeleno), vapnenci i dolomiti (narančasto), dolomiti (tamno crveno), crveni klasititi (tamno rozo), konglomerati i klastiti (rozo), dolomitizirani vapnenci (plavo), brečoliki vapnenci (svijetlo plavo), rudistni vapnenci (tirkizno), foraminiferski vapnenci (svijetlo zeleno), lapori (ljubičasto), lapori i pješčenjaci (tamno plavo), breče (zeleno), glaciofluvijalni sedimenti (crno), pijesak i šljunak (crveno), pijesci (žuto). Nijedan model nije bolji ili lošiji od drugog, već ovisi o svojstvima podataka i rezultatu koji se želi dobiti. Svaki model daje određeni tip podatka koji je karakterističan za njega u skladu s njegovim naredbama. Model koji više odgovara željenom rezultatu je model *rf*, koji se uzima za daljnja istraživanja (Slika 7.1.) i (Slika 7.2.). Sljedeći korak u istraživanju je detaljni vizualni pregled preko QGIS-a i dodatna provjera rezultata usporedbom OGK karata te izlaskom na teren.

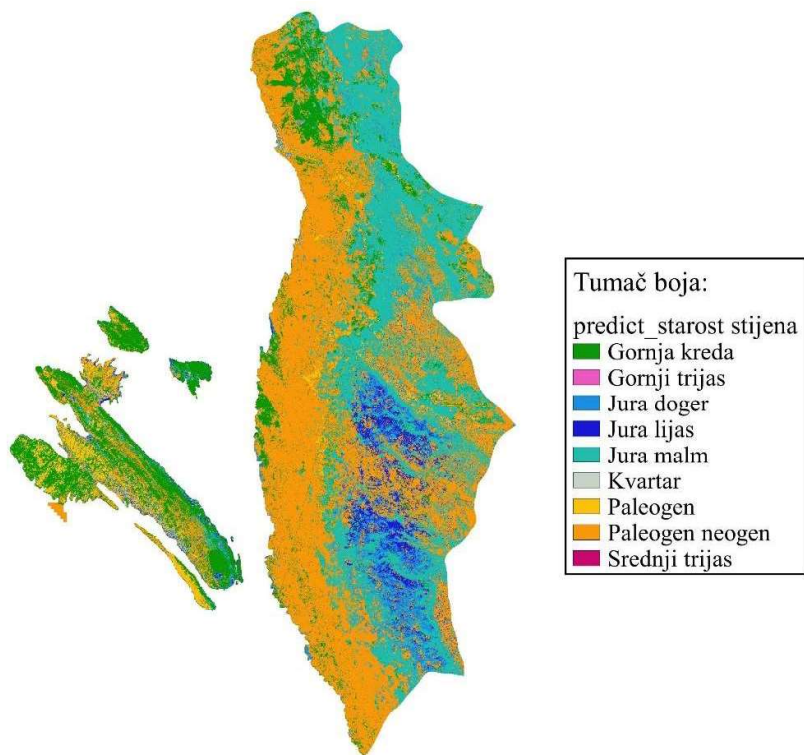


Slika 7.1. Predict\_rf starost stijena.

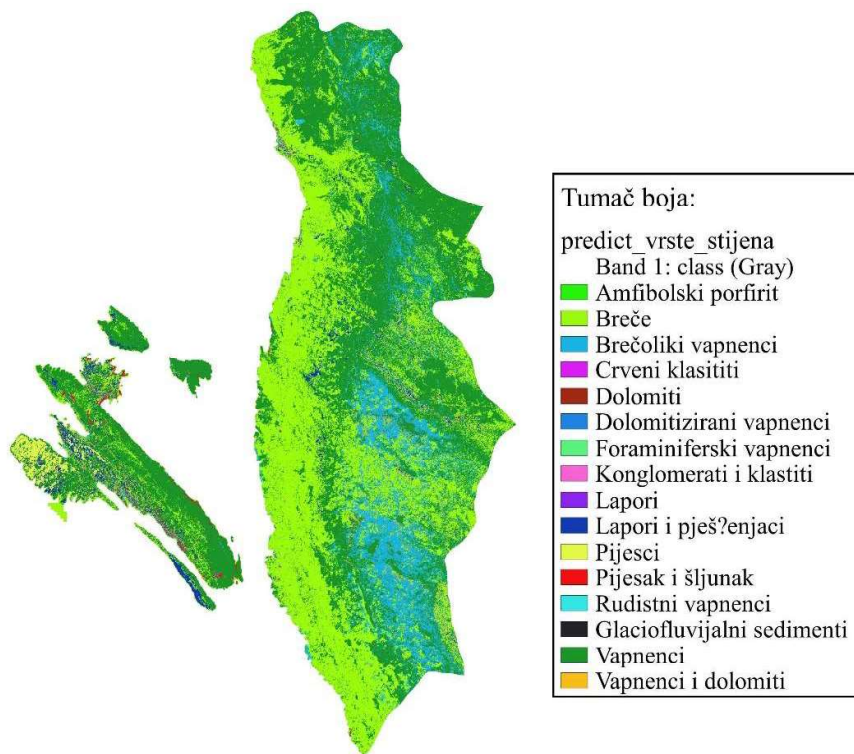


Slika 7.2. Predict\_rf vrste stijena.

Metoda *svm* se obično koristi za velike dimenzionalnost, oponašanje i prilagodbu bilo kojeg oblika za svaku zamišljenu klasifikaciju, ali za ovaj slučaj istraživanja rezultat nije zadovoljavajući (Slika 7.3.) i (Slika7.4.). Za razliku od RF-a koji se obično koristi za klasifikaciju kada postoji puno šuma i nejasnoća u podacima, ne pruža dobre rezultate za jasnu granicu između klasa za zadani problem. Izjednačava različite klase, većinu stavlja u pojedine klase pa samim time smanjuje pouzdanost klasifikacije litologije.

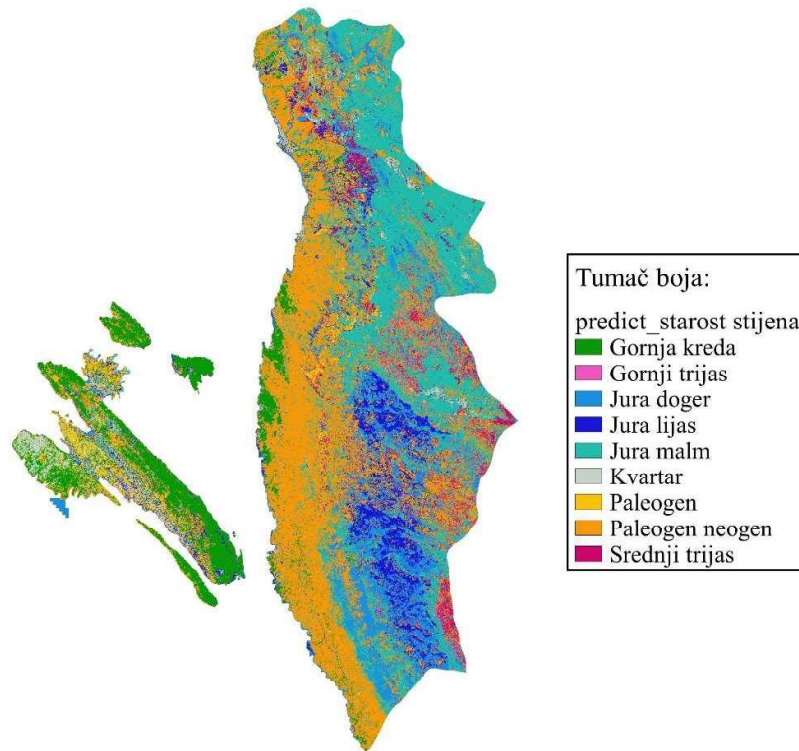


Slika 7.3. Predict\_svm starost stijena.

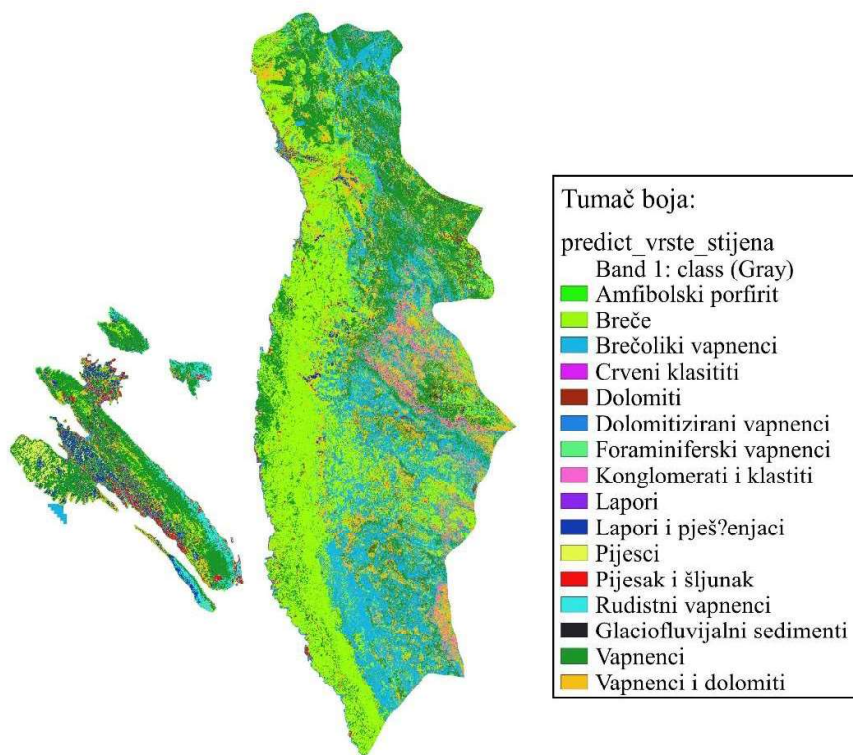


Slika 7.4. Predict\_svm vrste stijena.

Model *nnet* može zapaziti kompleksne i nepravilne oblike što ga čini snažnim alatom u klasifikaciji, u ovom slučaju rezultati su zadovoljavajući za zadani problem poslije metode RF-a (Slika 7.5.) i (Slika 7.6.). Za razliku od RF-a zahtijeva više vremena i veći broj parametara za trening. Jasno izdvaja većinu klasa i pouzdan je za daljnja istraživanja.



Slika 7.5. Predict\_nnet starost stijena.



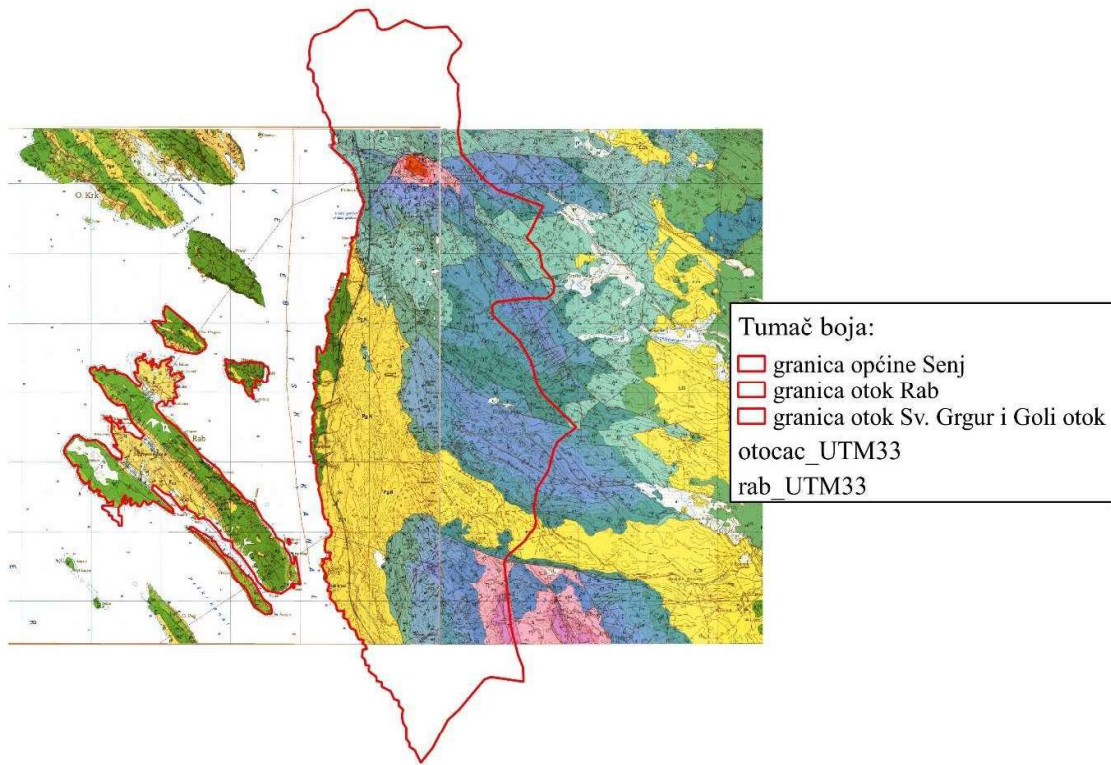
Slika 7.6. Predict\_nnet vrste stijena.

Model koji se uzima za daljnja istraživanja kao zadovoljavajući izbor za postizanje ciljanih rezultata u klasifikaciji litologije je *rf*. Iako se pokazalo da je model *nnet* također učinkovit, *rf* model ima svoje prednosti. Model *svm* nema jasno definirane granice što rezultira većom pogreškom u klasifikaciji i miješanju kanala te samim time se izostavlja za daljnja istraživanja.

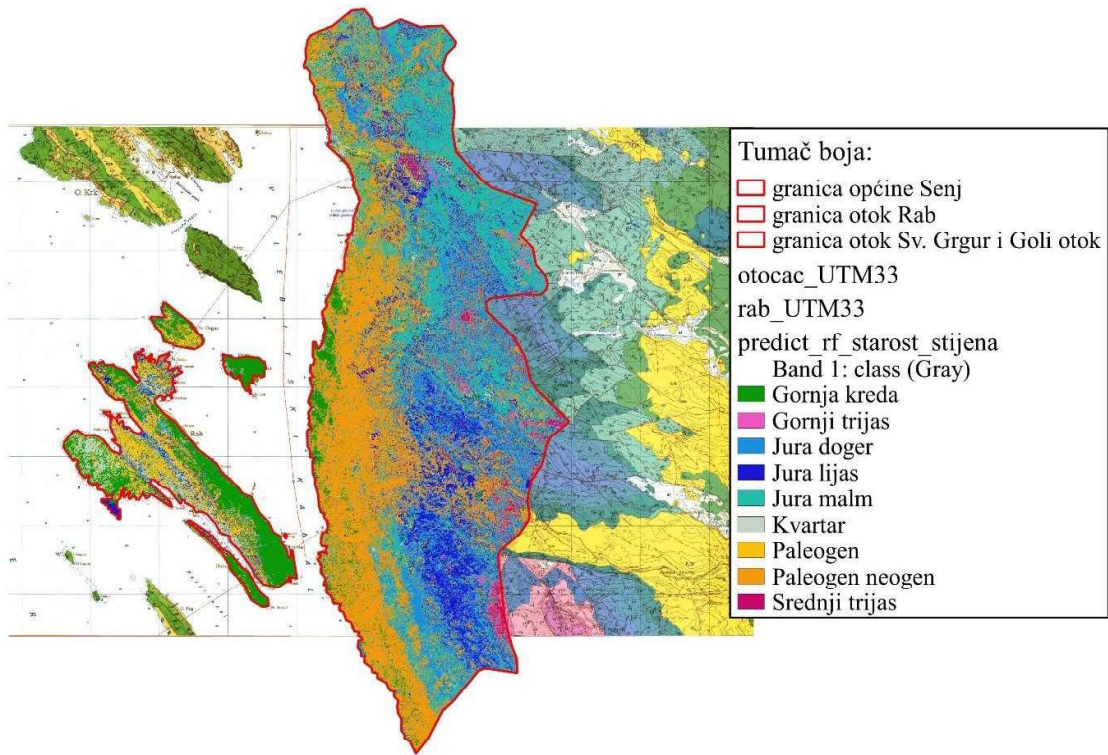


## 7.1. Potvrda modela *rf* starosti stijena sa kartama OGK Rab i Otočac

Primjenom vizualnog pregleda u QGIS-u dobiva se uvid u klasifikaciju litologije. Za jasniju usporedbu dobivenih rezultata devet grupa vremenskog razdoblja, boje su približno usklađene sa bojama koje se tradicionalno koriste na geološkim kartama. Srednji trijas (tamno rozo), gornji trijas (rozo), jura lijas (tamno plavo), jura doger (plavo), jura malm (tirkizno), gornja kreda (zeleno), paleogen (žuto), paleogen-neogen (narančasto), kvartar (sivo). Slika 7.1.1. prikazuje područje istraživanja na OGK Rabu i OGK Otočcu, dok slika 7.1.2. prikazuje klasifikaciju litologije modela *rf*.

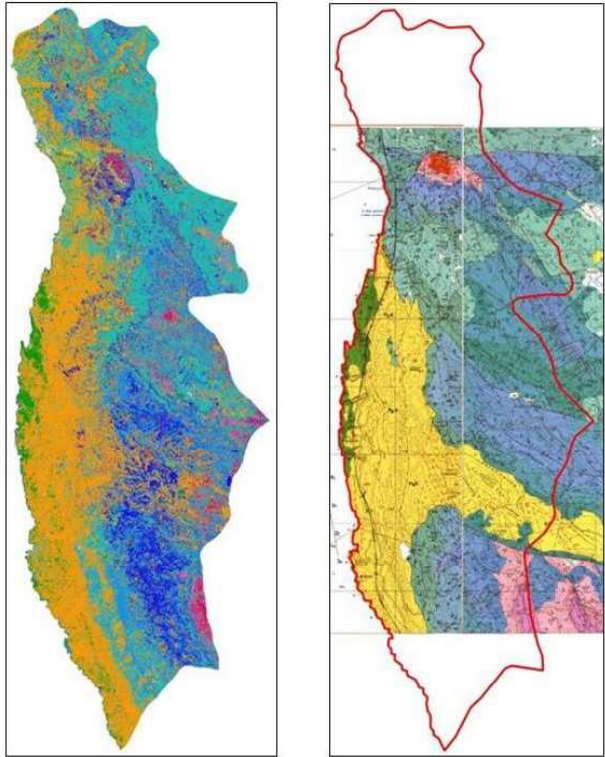


Slika 7.1.1. Područje istraživanja na OGK Rabu i OGK Otočcu.

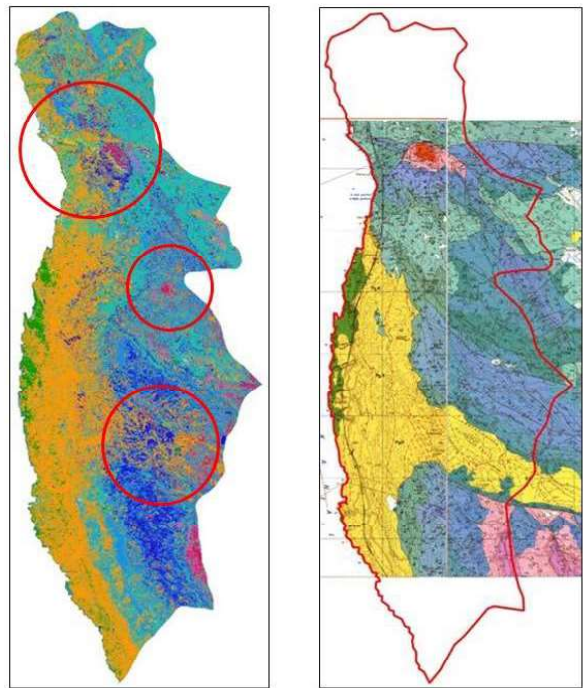


Slika 7.1.2. Klasifikacija litologije modela *rf*.

Usporedba modela *rf* i OGK Otočac za općinu Senj prikazano je na slici 7.1.3. U okviru analize geološke litologije provedeno je uspoređivanje dobivenog modela *randomForest* s Osnovnom geološkom kartom Otočac, kako bi se procijenila sposobnost algoritma za detekciju i kartiranje različitih geoloških starosti. Rezultati ove usporedbe ukazuju na nekoliko značajnih zapažanja, primjećuju se preklapanja geološke starosti između *randomForest* modela i Geološke karte. Uočena je bolja klasifikacija litologije na krškim područjima. Pretpostavlja se da je ovako dobra detekcija rezultat manje vegetacije na tim lokacijama, što omogućuje modelu bolju identifikaciju terenskih karakteristika. Visoka usklađenost između modela *rf* i OGK Otočac prisutna je kod gornje krede, označene zelenom bojom. Geološke starosti donje, srednje i gornje jure, označene tamno plavom, plavom i tirkiznom bojom, također su jasno vidljive i dobro iscrtnane. Kvartar, označen sivom bojom, detektiran je samo na manjim područjima, što ukazuje na izazove u identifikaciji ovog geološkog razdoblja. Paleogen-neogen, označen narančastom bojom, dobro je određen i iscrtnan u srednjem dijelu područja, no primjećuje se gubitak u desnom dijelu i višak identificiran u gornjem dijelu (Slika 7.1.4.). Geološke starosti gornjeg i srednjeg trijas, označene rozom i tamno rozom bojom, pokazuju sličnu situaciju kao paleogen-neogen, gdje su donji i gornji dijelovi dobro iscrtnani i detektirani, dok se u srednjem dijelu primjećuje višak (Slika 7.1.4.).

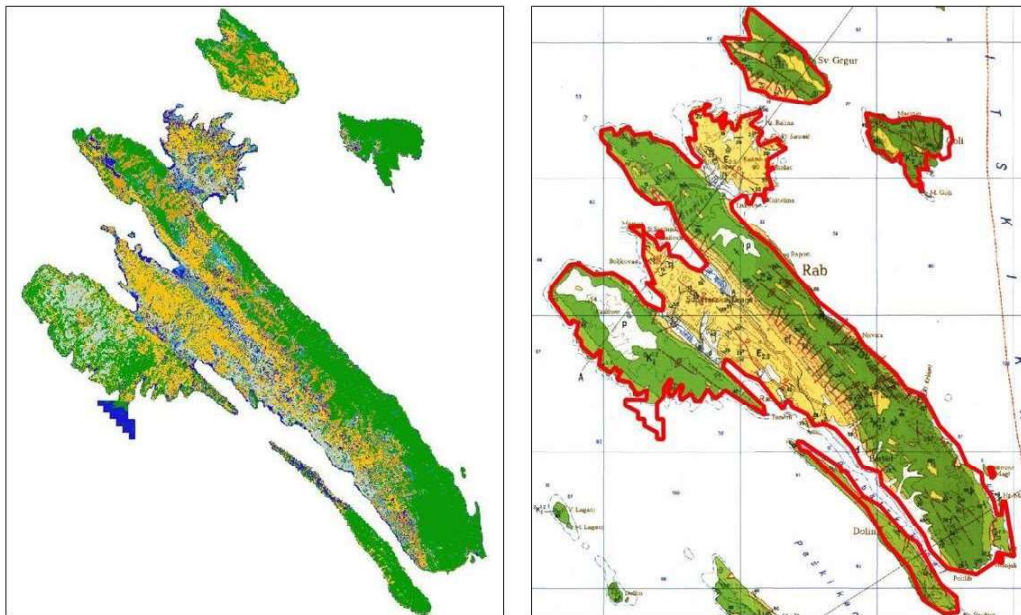


Slika 7.1.3. Usporedba litologije modela *rf* i OGK Otočac.

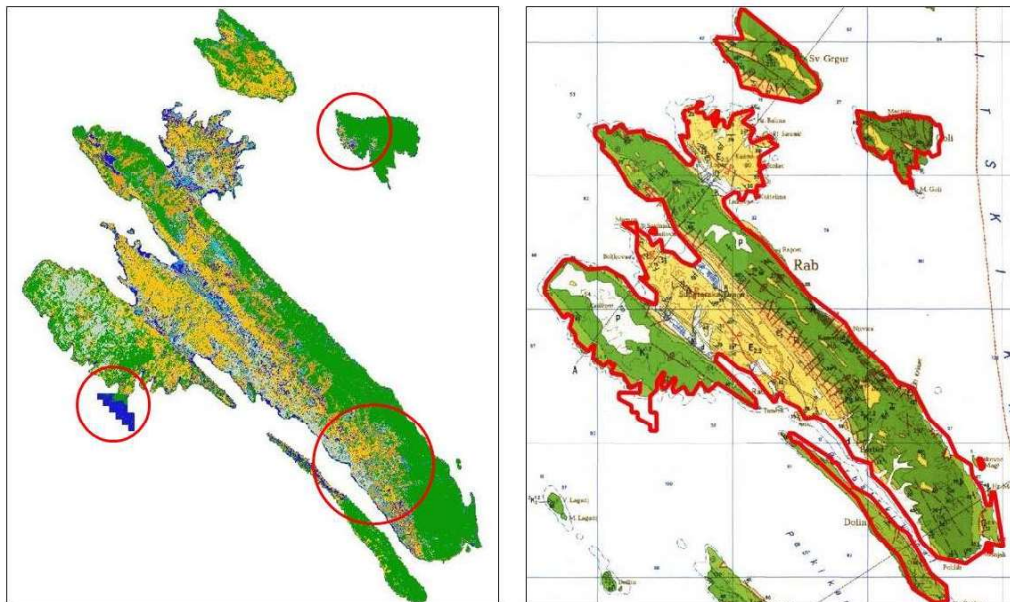


Slika 7.1.4. Razlika litologije modela *rf* naspram OGK Otočac

Usporedba modela *rf* i OGK Rab za otok Rab, Sv. Grgur i Goli otok prikazano je na slici 7.1.5. U okviru analize geološke litologije provedeno je uspoređivanje dobivenog modela Random Forest s Osnovnom geološkom kartom Rab. Visoka usklađenost između modela *rf* i OGK Rab prisutna je kod gornje krede, označene zelenom bojom. Geološka starost paleogena, označena žutom bojom, jasno je vidljiva i dobro određena u dijelovima na kojima se očekuje, iako se primjećuje mali gubitak na Golom otoku i višak u donjem dijelu otoka Raba (Slika 7.1.6.). Kvartar, označen sivom bojom, dobro je određen i iscrtan, no uočeno je miješanje sa geološkim razdobljem jure, označeno plavom bojom (Slika 7.1.6.). Za neispravnu detekciju na određenim područjima modela *rf*, pretpostavlja se kao rezultat veće gustoće vegetacije, manje izraženih krških karakteristika te odaziv sličnih frekvencija koje sustav stavlja u jednu skupinu, što dovodi sustav do izazova u određivanju geološke starosti.



Slika 7.1.5. Usporedba litologije modela *rf* i OGK Rab.



Slika 7.1.6. Razlika litologije modela *rf* naspram OGK Rab

Unatoč tome što je *randomForest* model u globalu uspješno obavio zadatak detekcije geološke litologije, primijećeno je stvaranje viška i miješanje pojedinih grupa, ukazujući na potrebu za daljnjim istraživanjima i razvojem naprednijih tehnologija za preciznu detekciju litologije na različitim geografskim područjima.

## 8. Zaključak

Korištenjem slobodno dostupnih multispektralnih snimaka iz satelitske misije Sentinel-2 (Europske Svemirske Agencije, Copernicus) i upotrebom aplikacije i programskog jezika otvorenog koda (QGIS i R programski jezik) stvorio se novi model za klasifikaciju litologije. Model za predikciju litologije stvoren je pomoću algoritama strojnog učenja. Primjenjujući program QGIS izradili su se poligoni starosti i vrste stijena, koji su služili kao podaci za uvježbavanje sustava strojnog učenja. Korištenjem algoritama nadzirane klasifikacije strojnog učenja *randomForest*, *Support Vector Machine* i *Neural Network* iz paketa CARET stvoreni su novi modeli za detekciju litologije. Uz samu detekciju litologije odvajale su se klase geološke starosti poput srednjeg i gornjeg trijasa, donje, srednje i gornje jure, gornje krede, paleogena, plaeogena-neogena i kvartara, te vrste stijena za pojedina razdoblja. Izradene su dvije detaljne analize spektralnog potpisa satelitske misije Sentinel-2. Prva spektralna analiza usmjerena je na određivanje starosti, a druga na identifikaciju različitih vrsta stijena u promatranom području. U obje analize, koristili su se podaci dobiveni kroz deset različitih kanala. Novostvorenim modelima predikcije izvela se nova detekcija litologije na multispektralnim satelitskim snimkama Sentinel-2 misije i dobila su se tri klasificirana rastera. Grafičkim prikazom tih triju klasificiranih rastera u QGIS-u moglo se vidjeti i koji od njih je pouzdaniji u predikciji litologije. Model koji se uzeo za daljnja istraživanja kao zadovoljavajući izbor za postizanje ciljanih rezultata u klasifikaciji litologije je *randomForest*. Primjenom vizualnog pregleda u QGIS-u dobiva se uvid u klasifikaciju litologije. Za jasniju usporedbu dobivenih rezultata devet grupa geološkog razdoblja, boje su približno usklađene sa bojama koje se tradicionalno koriste na geološkim kartama. U okviru analize geološke litologije provedeno je uspoređivanje dobivenog modela *randomForest* s Osnovnom geološkom kartom, kako bi se procijenila sposobnost algoritma za detekciju i kartiranje različitih geoloških starosti. Rezultati ove usporedbe ukazuju na nekoliko značajnih zapažanja, primjećuju se preklapanja geološke starosti, uočena je bolja klasifikacija litologije na krškim područjima. Unatoč tome što je *randomForest* model u globalu uspješno obavio zadatak detekcije geološke litologije, primijećeno je stvaranje viška i miješanje pojedinih grupa. Neispravna detekcija na određenim područjima, rezultat je veće gustoće vegetacije, manje izraženih krških karakteristika te odaziv sličnih frekvencija koje sustav stavlja u jednu skupinu. Rezultati ovog istraživanja pokazuju da metodama daljinskog istraživanja i korištenjem slobodno dostupnih multispektralnih snimaka Sentinel-2 kao i metoda strojnog učenja, moguće izraditi model za klasifikaciju litologije. Kako bi se poboljšala efikasnost ovog modela potrebni su novi podaci s većom prostornom rezolucijom kako bi se povećala pouzdanost detekcije. Osim podataka s većom prostornom rezolucijom trebali bi se koristiti i satelitski sintetički radarski snimci iz satelitske misije Sentinel-1. Daljnja istraživanja trebala bi se usmjeriti za pronalazak naprednije tehnologije za potrebe klasifikacije litologije na različitim područjima istraživanja, koja sadrži informacije o starosti i vrsti stijena, čime inženjerima geologije omogućava brže i ekonomičnije istraživanje i razumijevanje geoloških karakteristika terena.

## **9. Zahvale**

Želim izraziti svoju zahvalnost svima koji su mi pomogli i podržali tijekom izrade ovog rada. Prije svega, želimo se posebno zahvaliti mojem mentoru doc. dr. sc. Ivanu Medvedu na neizmjernej pomoći, vodstvu, odvojenom vremenu i usmjeravanju koje je pružio tijekom cijelog rada. Bez njegove stručnosti, iskustva, entuzijazma i savjeta, ovaj rad ne bi bio ovako kvalitetan i značajan. Hvala Vam na Vašem strpljenju i podršci koju ste pružali u svakom trenutku.

Želim se zahvaliti obitelji koja me je podržavala tijekom cijelog rada.

Svima Vam još jednom želim izraziti iskrenu zahvalnost i poštovanje.

## 10. Literatura

Abd El-Wahed, M., Kamh, S., Abu Anbar, M., Zoheir, B., Hamdy, M., Abdeldayem, A., Lebda, E.M., Attia, M. (2023). Multisensor Satellite Dana and Field Studies for Unravelling the Structural Evolution and Gold Metallogeny of the Gerf Ophiolitic Nappe, Eastern Desert, Egypt. *Remote Sens.* 15, 1974.

Basu, B., Sannigrahi, S., Sarkar Basu, A., & Pilla, F. (2021). Development of novel classification algorithms for detection of floating plastic debris in coastal waterbodies using multispectral Sentinel-2 remote sensing imagery. *Remote Sensing*, 13(8), 1598.

Brcković, A. (2022). Granulometrijska heterogenost pliokvartarnih naslaga kao važan čimbenik aktivacije klizišta. Diplomski rad Zagreb: Rudarsko-geološko-naftni fakultet.

Dobrinčić, D., Gašparović, M., & Medak, D. (2021). Sentinel-1 and 2 time-series for vegetation mapping using random forest classification: A case study of Northern Croatia. *Remote Sensing*, 13(12), 2321.

Giorgi, F. M., Ceraolo, C., & Mercatelli, D. (2022). The R language: an engine for bioinformatics and data science. *Life*, 12(5), 648.

Jukić, M. (2021). Izdavanje litoloških jedinica multispektralnom analizom na području Sjevernog Velebita. Diplomski rad Zagreb: Rudarsko-geološko-naftni fakultet.

Komac, M., Ribičić, M. (1998). The application of remote sensing - satellite imagery in engineering geology (study area of Črni Kal, Slovenia). *Geologija* 41, 411-434.

Lemenkova, P. (2020). Using R Packages' tmap', raster'and'ggmap'for Dartographic Visualization: An Example of Dem-Based Terrain Modelling of Italy, Apennine Peninsula. *Zbornik radova-Geografski fakultet Univerziteta u Beogradu*, 68, 99-116.

Mamužić, P. (1962). Novija geološka istraživanja otoka Raba. *Geološki vjesnik Instituta za geološka istraživanja u Zagreb i Hrvatskog geološkog društva*, 15(1), 121-142.

Milan, A. (1970). Prilog poznavanja geološke građe šire okolice Senja. *Senjski zbornik: prilozi za geografiju, etnologiju, gospodarstvo, povijest i kulturu*, 4(1), 33-44.

Zhang, T., Su, J., Xu, Z., Luo, Y., & Li, J. (2021). Sentinel-2 satellite imagery for urban land cover classification by optimized random forest classifier. *Applied Sciences*, 11(2), 543.

### WEB IZVORI:

Copernicus-a, 2023. *About Copernicus*. URL: <https://www.copernicus.eu/en/about-copernicus> (1.3.2023.)

Copernicus-b, 2023. Copernicus Open Access Hub. URL: <https://scihub.copernicus.eu/dhus/#/home> (1.3.2023.)

CRAN-a, 2023. *Leafsync: Small Multiples for Leaflet Web Maps*. URL: <https://cran.r-project.org/web/packages/leafsync/index.html> (31.3.2023.)

CRAN-b, 2023. *Mapview: Interactive Viewing of Spatial Dana in R*. URL: <https://cran.r-project.org/web/packages/mapview/index.html> (31.3.2023.)

CRAN-c, 2023. *NeuralNetTools: Visualisation and Analysis Tools for Neural Networks*. URL: <https://cran.r-project.org/web/packages/NeuralNetTools/index.html>. (31.3.2023.)



CRAN-d, 2023. *RColorBrewer: ColorBrewer Palettes*, URL: <https://cran.r-project.org/web/packages/RColorBrewer/index.html> (31.3.2023.)

CRAN-e, 2023. *LiblineaR: Linear Predictive Models Based on the LIBLINEAR C/C++ Library*. URL: <https://cran.r-project.org/web/packages/LiblineaR/index.html> (31.3.2023.)

CRAN-f, 2023. *Terra: Spatial Data Analysis*. URL: <https://cran.r-project.org/web/packages/terra/index.html> (31.3.2023.)

Datacamp, 2023. *Building Neural Network (NN) Models in R*. URL: <https://www.datacamp.com/tutorial/neural-network-models-r> (2.4.2023.)

Datacamp, 2018. *Support Vector Machines in R*. URL: <https://www.datacamp.com/tutorial/support-vector-machines-r> (2.4.2023.)

Geeksforgeeks, 2021. *Classifying data using Support Vector Machines (SVMs) in R*. URL: <https://www.geeksforgeeks.org/classifying-data-using-support-vector-machines-svms-in-r/> (3.4.2023.)

Geeksforgeeks, 2020. *How Neural Networks are used for Classification in R Programming*. URL: <https://www.geeksforgeeks.org/how-neural-networks-are-used-for-classification-in-r-programming/> (3.4.2023.)

GISGeography, 2022. *Sentinel 2 Bands and Combinations*. URL: <https://gisgeography.com/sentinel-2-bands-combinations/> (4.3.2023.)

GitHub, 2023. *MLmetrics*. URL: <https://github.com/yanyachen/MLmetrics> (1.4.2023.)

Javatpoint-a, 2021. *List of R packages*. URL: <https://www.javatpoint.com/list-of-r-packages> (1.4.2023.)

Javatpoint-b, 2021. *Random Forest Algorithm*. URL: <https://www.javatpoint.com/machine-learning-random-forest-algorithm> (1.4.2023.)

Kent State University, 2021. *STATISTICAL&QUALITATIVE DANA ANALYSIS SOFTWARE: ABOUT RAND RSTUDIO*. URL: <https://libguides.library.kent.edu/statconsulting/r> (28.2.2023.)

MedMPAnet projekt, 2014. *Kartiranje morskih i obalnih staništa i vrsta otoka Prvić, Goli otok, Sv. Grgur, Unije, Susak i Srakane*. URL: [https://www.rac-spa.org/sites/default/files/doc\\_medmpanet/final\\_docs\\_croatia/15\\_mapping\\_marine\\_and\\_coastal\\_croatia.pdf](https://www.rac-spa.org/sites/default/files/doc_medmpanet/final_docs_croatia/15_mapping_marine_and_coastal_croatia.pdf) (20.9.2023.)

Packt, 2023. *Pros and cons od neural network*. URL: <https://subscription.packtpub.com/book/big-data-&-business-intelligence/9781788397872/1/ch011v11sec27/pros-and-cons-of-neural-networks> (2.4.2023.)

Patwal, M., 2020. *A Short Introduction to R Programming*. URL: <https://bookdown.org/manishpatwal/bookdown-demo/basic-operations-in-r.html> (1.3.2023.)

Posit Support, 2023. *Recommended Packages*. URL: <https://support.posit.co/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages> (1.4.2023.)

QGIS-a, 2023. *Features*. URL: [https://docs.qgis.org/3.28/en/docs/user\\_manual/preamble/features.html](https://docs.qgis.org/3.28/en/docs/user_manual/preamble/features.html) (1.3.2023.)

QGIS-b, 2023. *About QGIS*. URL: <https://www.qgis.org/en/site/about/index.html> (1.3.2023.)

Rdrr, 2023. *The R Graphics Devices and Support for Colours and Fonts*. URL: <https://rdrr.io/r/grDevices/grDevices-package.html> (1.4.2023.)

R Learning Infrastructure Home Web Page, 2023. *Introduction to R and RStudio*. URL: [http://mercury.webster.edu/aleshunna/R\\_learning\\_infrastructure/Introduction\\_to\\_R\\_and\\_RStudio.html](http://mercury.webster.edu/aleshunna/R_learning_infrastructure/Introduction_to_R_and_RStudio.html) (1.3.2023.)

R package Wikipedia, 2023. URL: [https://en.wikipedia.org/wiki/R\\_package](https://en.wikipedia.org/wiki/R_package) (1.4.2023.)

RStudio Wikipedia, 2023. URL: <https://en.wikipedia.org/wiki/RStudio> (28.2.2023.)

Sentinel Online, 2023. *SENTINEL-2 MISSION GUIDE*. URL: <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2> (4.3.2023.)

Simplilearn, 2023. *Introduction to Random Forest in R*. URL: <https://www.simplilearn.com/tutorials/data-science-tutorial/random-forest-in-r> (30.3.2023.)

Spatial Data Science, 2023. *The terra package*. URL: <https://rspatial.org/pkg/1-introduction.html#the-terra-package> (1.4.2023.)

Tutorialspoint, 2023. *R-Packages*. URL: [https://www.tutorialspoint.com/r/r\\_packages.htm](https://www.tutorialspoint.com/r/r_packages.htm) (3.4.2023.)

UC Business Analytics R Programming Guide, 2023. *Support Vector Machine*. URL: <https://uc-r.github.io/svm> (3.4.2023.)