

# Određivanje poplavljenog područja na području Karlovačke i Sisačko-moslavačke županije korištenjem multispektralnih satelitskih snimaka Sentinel-2 u programskom jeziku R

---

Drešar, Iva

Master's thesis / Diplomski rad

2024

*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://um.nsk.hr/um:nbn:hr:169:523953>

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

*Download date / Datum preuzimanja:* **2025-03-14**



*Repository / Repozitorij:*

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



SVEUČILIŠTE U ZAGREBU  
RUDARSKO-GEOLOŠKO-NAFTNI FAKULTET  
Diplomski studij geološkog inženjerstva

**ODREĐIVANJE POPLAVLJENOG PODRUČJA NA PODRUČJU KARLOVAČKE  
I SISAČKO-MOSLAVAČKE ŽUPANIJE KORIŠTENJEM MULTISPEKTRALNIH  
SATELITSKIH SNIMAKA SENTINEL-2 U PROGRAMSKOM JEZIKU R**

Diplomski rad

Iva Drešar

G2200

Zagreb, 2024.



KLASA: 602-02/24-01/16  
URBROJ: 251-70-13-24-1  
U Zagrebu, 12.09.2024.

Iva Drešar, studentica

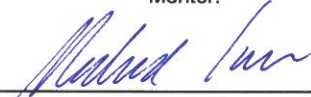
## RJEŠENJE O ODOBRENJU TEME

Na temelju vašeg zahtjeva primljenog pod KLASOM 602-02/24-01/16, URBROJ: 251-70-13-24-1 od 19.06.2024. priopćujemo vam temu diplomskog rada koja glasi:

### ODREĐIVANJE POPLAVLJENOG PODRUČJA NA PODRUČJU KARLOVAČKE I SISAČKO-MOSLAVAČKE ŽUPANIJE KORIŠTENJEM MULTISPEKTRALNIH SATELITSKIH SNIMAKA 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)

## **ZAHVALA**

*Prije svega, želim se zahvaliti svom mentoru doc. dr. sc. Ivan Medvedu na strpljenju, podršci i dragocjenim savjetima koje mi je pružio tokom cijelog procesa pisanja ovog diplomskog rada.*

*Zahvaljujem se i svojim profesorima i asistentima Rudarsko-geološko-naftnog fakulteta na prenesenom znanju i motivaciji tokom cijelog studiranja.*

*Posebnu zahvalnost dugujem svojoj obitelji i prijateljima. Hvala vam na neizmjernej podršci, razumijevanju i ohrabrivanju u trenucima kada je bilo najpotrebnije.*

*Na kraju, zahvaljujem se svim kolegama studentima, sa kojima sam dijelila radosti i izazove studiranja. Vaša podrška i druženje učinili su ovaj period mog života lijepim.*

ODREĐIVANJE POPLAVLJENOG PODRUČJA NA PODRUČJU KARLOVAČKE I SISAČKO-  
MOSLAVAČKE ŽUPANIJE KORIŠTENJEM MULTISPEKTRALNIH SATELITSKIH SNIMAKA  
SENTINEL-2 U PROGRAMSKOM JEZIKU R

Iva Drešar

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

Sažetak

Klasifikacija terena poplavljenog područja Karlovačke i Sisačko-moslavačke županije korištenjem multispektralnih satelitskih snimaka Sentinel-2 prije i poslije poplave. Satelitske snimke preuzete sa službene stranice geoportala Copernicus. Osim satelitskih snimaka, ulazni podaci su područje istraživanja i poligoni za treniranje izrađeni u programu ArcMap. Klasifikacija terena izvedena je strojnim učenjem korištenjem algoritama Random Forest, Support Vector Machine i Neural Network u programskom jeziku R.

Ključne riječi: Poplavljena područja, klasifikacija, predikcija, Sentinel-2, RF, SVM, NNET, ArcMap, R, NDWI

Završni rad sadrži: 60 stranice, 19 tablica, 52 slika, 0 priloga, i 34 reference.

Jezik izvornika: Hrvatski

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

Mentori: Dr. sc. Ivan Medved, docent RGNF

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

DETERMINATION OF THE FLOODED AREA IN THE AREA OF KARLOVAČKA AND SISAČKO-  
MOSLAVACKA COUNTIES USING MULTISPECTRAL SATELLITE IMAGES SENTINEL-2 IN  
SOFTWARE LANGUAGE R

Iva Drešar

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

Classification of flood-affected areas in the Karlovac and Sisak-Moslavina counties using Sentinel-2 multispectral satellite images before and after the flood. The satellite images were downloaded from the official Copernicus geoportal website. In addition to the satellite images, the input data include the study area and training polygons created in ArcMap. The terrain classification was performed using machine learning algorithms, including Random Forest, Support Vector Machine, and Neural Network, in the R programming language.

Keywords: Flooded areas, classification, prediction, Sentinel-2, RF, SVM, NNET, ArcMap, R, NDWI

Thesis contains: 60 pages, 19 tables, 52 figures, 0 appendixes, i 34 references.

Original in: English

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

Supervisors: Assistant Professor Ivan Medved, PhD

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

## SADRŽAJ

1. UVOD	1
2. POPLAVLJENA PODRUČJA KARLOVAČKE I SISAČKO-MOSLAVAČKE ŽUPANIJE	2
2.1. Geologija	3
2.2. Klimatski uvjeti	3
3. PODRUČJE ISTRAŽIVANJA	4
4. TEORIJSKE OSNOVE	5
4.1. Copernicus (ESA)	5
4.1.1. Sentinel-2	6
4.1.2. Geoportal Copernicus	8
4.2. ArcGIS (GIS) – ESRI	9
4.3.1. ArcMap	9
4.3. Programski jezik R	11
4.2.1. Sučelje R Studio	11
4.2.2. R paketi	13
4.2.1. Metode strojnog učenja („Machine learning“)	16
5. PRIPREMA PODATAKA	19
6. OBRADA PODATAKA	23
6.1. Metoda Random Forest (RF)	32
6.2. Metoda Support Vector Machine (SVM)	37
6.3. Metoda Neural Network (NNET)	40
6.4. Usporedba modela RF, SVM i NNET	44
7. ANALIZA REZULTATA	47
8. ZAKLJUČAK	55
9. LITERATURA	57

## POPIS SLIKA

<b>Slika 2.1.</b> Poplava na području Karlovca (Telegram, 2023)	2
<b>Slika 2.2.</b> Isječak Osnovne geološke karte (OGK), list Sisak (M. Pikija i dr., 1987). Originalno mjerilo OGK je 1:100.000.	3
<b>Slika 2.3.</b> Osnovna geološka karta, list Karlovac (J. Bukovac, N. Magaš, A. Šimunić i dr., 2014)	2
<b>Slika 2.4.</b> Osnovna geološka karta, list Sisak (M. Pikija i dr., 1987).	2
<b>Slika 3.1.</b> Rijeka Kupa od Karlovca do Siska i označeno područje istraživanja plavim kvadratom	4
<b>Slika 3.2.</b> Područje istraživanja	4
<b>Slika 4.1.</b> Misije promatranja Zemlje koje je razvila ESA (Copernicus-2, 2024).	5
<b>Slika 4.2.</b> Raspored dvaju satelita Sentinel-2 (Sentinel-a, 2024).	6
<b>Slika 4.3.</b> Sučelje Copernicus Browser-a (Copernicus Browser-2, 2024).	8
<b>Slika 4.4.</b> Sučelje programa ArcMap	9
<b>Slika 4.5.</b> Razlika između rasterskih i vektorskih podataka (GIS, 2024)	10
<b>Slika 4.6.</b> Vektorski i rasterski podaci različitog sadržaja u obliku slojeva (City of Newberg, 2024)	10
<b>Slika 4.7.</b> Sučelje R Studio-a	12
<b>Slika 4.8.</b> Način rada algoritma Random Forest (Javapoint, 2021)	16
<b>Slika 4.9.</b> Način rada algoritma Support Vector Machine (Datacamp, 2018)	17
<b>Slika 4.10.</b> Način rada algoritma Neural Network (Geeksforgeeks, 2023)	18
<b>Slika 5.1.</b> Odabir satelitske misije, maksimalne pokrivenosti naoblakom i vremenskog raspona (Copernicus Browser-2, 2024).	19
<b>Slika 5.2.</b> Multispektralne satelitske snimke prije poplave (18.3.2023.) (Copernicus Browser-2, 2024)	20
<b>Slika 5.3.</b> Multispektralne satelitske snimke poslije poplave (22.5.2023.) (Copernicus Browser-2, 2024)	20
<b>Slika 5.4.</b> Područje istraživanja definirano u programu ArcMap	21
<b>Slika 5.5.</b> Poligoni za strojno učenje definirani u programu ArcMap	21
<b>Slika 5.6.</b> Atributna tablica sa definiranim klasama za svaki poligon	22
<b>Slika 6.1.</b> Folder R20m sa kanalima za terensku klasifikaciju	24
<b>Slika 6.2.</b> Histogrami za svaki kanal multispektralnih snimaka – prije poplave.	29
<b>Slika 6.3.</b> Histogrami za svaki kanal multispektralnih snimaka – poslije poplave	29



<b>Slika 6.4.</b> Grafički prikaz modela RF – prije i poslije poplave _____	32
<b>Slika 6.5.</b> Prikaz grafikona heatmap (1) – prije i poslije poplave _____	35
<b>Slika 6.6.</b> Prikaz grafikona heatmap (2) – prije i poslije poplave _____	35
<b>Slika 6.7.</b> Grafikon za brzu procjenu ključnih varijabli – prije poplave _____	36
<b>Slika 6.8.</b> Grafikon za brzu procjenu ključnih varijabli – poslije poplave _____	36
<b>Slika 6.9.</b> Grafički prikaz modela SVM – prije i poslije poplave _____	38
<b>Slika 6.10.</b> Grafički prikaz modela NNET – prije i poslije poplave _____	40
<b>Slika 6.11.</b> Graf neuronske mreže – prije i poslije poplave _____	43
<b>Slika 6.12.</b> Neuronska mreža – prije poplave _____	43
<b>Slika 6.13.</b> Neuronska mreža – poslije poplave _____	43
<b>Slika 6.14.</b> Grafikon za usporedbu modela RF, SVM i NNET – prije poplave _____	45
<b>Slika 6.15.</b> Grafikon za usporedbu modela RF, SVM i NNET – poslije poplave _____	45
<b>Slika 6.16.</b> Predikcije modela RF, SVM i NNET za satelitske snimke prije poplave _____	46
<b>Slika 6.17.</b> Predikcije modela RF, SVM i NNET za satelitske snimke poslije poplave _____	46
<b>Slika 7.1.</b> Digitalna ortofoto karta (DOF) preklapljena sa predikcijom _____	47
<b>Slika 7.2.</b> Satelitska snimka preklapljena sa i bez predikcije _____	48
<b>Slika 7.3.</b> Predikcija pomoću modela RF – prije poplave _____	49
<b>Slika 7.4.</b> Predikcija pomoću modela SVM – prije poplave _____	49
<b>Slika 7.5.</b> Predikcija pomoću modela NNET – prije poplave _____	49
<b>Slika 7.6.</b> Predikcija pomoću modela RF – poslije poplave _____	50
<b>Slika 7.7.</b> Predikcija pomoću modela SVM – poslije poplave _____	50
<b>Slika 7.8.</b> Predikcija pomoću modela NNET – poslije poplave _____	51
<b>Slika 7.9.</b> Predikcija modela RF preklapljena sa NDWI indeksom _____	52
<b>Slika 7.10.</b> Predikcija modela SVM preklapljena sa NDWI indeksom _____	52
<b>Slika 7.11.</b> Predikcija modela NNET preklapljena sa NDWI indeksom _____	52
<b>Slika 7.12.</b> Satelitska snimka poslije poplave _____	53
<b>Slika 7.13.</b> Usporedba NDWI indeksa, predikcija modela RF, SVM, NET i satelitske snimke poslije poplave _____	54

## POPIS TABLICA

<b>Tablica 2.1.</b> Maksimalni vodostaji i protoci na hidrološkim postajama rijeke Korane, Kupe i Une (DHMZ-b, 2024)	3
<b>Tablica 4.1.</b> Sentinel-2 kanali (Sentinel-a, 2024).	7
<b>Tablica 4.2.</b> Korišteni paketi u R programskom jeziku (posit <sup>TM</sup> Support, 2024; RDocumentation, 2024)	14
<b>Tablica 6.1.</b> Prikaz vrijednosti piksela iz različitih kanala s pridodanom klasom – prije poplave	27
<b>Tablica 6.2.</b> Prikaz vrijednosti piksela iz različitih kanala s pridodanom klasom – poslije poplave	28
<b>Tablica 6.3.</b> Podaci za treniranje – prije i poslije poplave	30
<b>Tablica 6.4.</b> Podaci za testiranje – prije i poslije poplave	30
<b>Tablica 6.5.</b> Usporedba definiranih i predikcijskih klasa modela RF – prije i poslije poplave	33
<b>Tablica 6.6.</b> Sveukupnu statistika modela RF – prije i poslije poplave	34
<b>Tablica 6.7.</b> Statistika modela RF – prije poplave	34
<b>Tablica 6.8.</b> Statistika modela RF – poslije poplave	34
<b>Tablica 6.9.</b> Usporedba definiranih i predikcijskih klasa modela SVM – prije i poslije poplave	38
<b>Tablica 6.10.</b> Sveukupnu statistika modela SVM – prije i poslije poplave	38
<b>Tablica 6.11.</b> Statistika modela SVM – prije poplave	39
<b>Tablica 6.12.</b> Statistika modela SVM – poslije poplave	39
<b>Tablica 6.13.</b> Usporedba definiranih i predikcijskih klasa modela NNET – prije i poslije poplave	41
<b>Tablica 6.14.</b> Sveukupnu statistika modela NNET – prije i poslije poplave	41
<b>Tablica 6.15.</b> Statistika modela NNET – prije poplave	41
<b>Tablica 6.16.</b> Statistika modela NNET – poslije poplave	42

## **POPIS KORIŠTNIH KRATICA**

<b>ArcGIS</b>	Geografski informacijski sustav (GIS) softver razvijen od strane tvrtke ESRI
<b>DHMZ</b>	Državni hidrometeorološki zavod
<b>ESA</b>	Europska svemirska agencija (European Space Agency)
<b>ESRI</b>	Environmental Systems Research Institute
<b>GIS</b>	Geografski informacijski sustav (Geographic Information System(s))
<b>NDWI</b>	Indeks normalizirane razlike vode (Normalized Difference Water Index)
<b>NNET</b>	Neuronska mreža (Neural Network)
<b>OLI</b>	Operation Land Imager
<b>OGK</b>	Osnovna geološka karta
<b>QGIS</b>	Geografski informacijski sustav otvorenog koda (Quantum Geographic Information System)
<b>R</b>	Programski jezik
<b>RF</b>	Slučajna šuma (Random Forest)
<b>RGB</b>	Red, green and blue
<b>L2A</b>	Sentinel-2 Level-2A
<b>SVM</b>	Support Vector Machine
<b>SWIR</b>	Kratkovalni infracrveni (Short Wave Infra-Red)
<b>TIFF</b>	Format datoteke za pohranu slika (Tagged Image File Format)
<b>VNIR</b>	Vidljivi i bliski infracrveni (Visible and Near-Infra-Red)

## 1. UVOD

Zadatak ovog diplomskog rada je klasifikacija terena na četiri klase: infrastruktura, poljoprivreda, šuma i vodene površine. Klasifikacija terena je izvedena korištenjem multispektralnih satelitskih snimaka i strojnim učenjem („Machine learning“) u programskom jeziku R. Multispektralne satelitske snimke, kao podloga za klasifikaciju terena, dobivene su pomoću Sentinel-2 satelita, a preuzete sa geoportala Copernicus koji se nalazi na službenoj ESA stranici. Također, korištenjem programa ArcMap definirano je područje istraživanja i poligoni koji predstavljaju ulazni podatak za strojno učenje u programskom jeziku R. Dakle, strojno učenje izvodi se pomoću programskog jezika R na način da se sustav prvo mora trenirati. Na temelju dobivenih podataka u obliku poligona i stečenog znanja sustav detektira sličnosti koje povlači između dobivenih i novih podataka. Na taj način dobiva se klasifikacija terena odabranog područja koje se istražuje. U sklopu istraživanja planirano je analizirati dva snimka, jedan prije i jedan poslije poplave, kako bi se precizno odredila poplavljena područja. Primarni cilj je identifikacija poplavljenih područja uz provođenje ostalih klasifikacija. Strojno učenje izvedeno je pomoću tri metode: *Random Forest (RF)*, *Support Vector Machine (SVM)* i *Neural Network (NNET)*. Algoritmi dobiveni tim metodama se na kraju uspoređuju te se izdvajaju predikcije koje daju relativno najbolje rezultate. Finalna tri rezultata dobivena programskim jezikom R su dodatno obrađena u programu ArcMap. Kao očekivani rezultat klasifikacijom terena pomoću strojnog učenja je samostalno strojno prepoznavanje poplavljenih područja kao i vidljiva razlika između satelitskih snimaka prije i poslije poplave.

## 2. POPLAVLJENA PODRUČJA KARLOVAČKE I SISAČKO-MOSLAVAČKE ŽUPANIJE

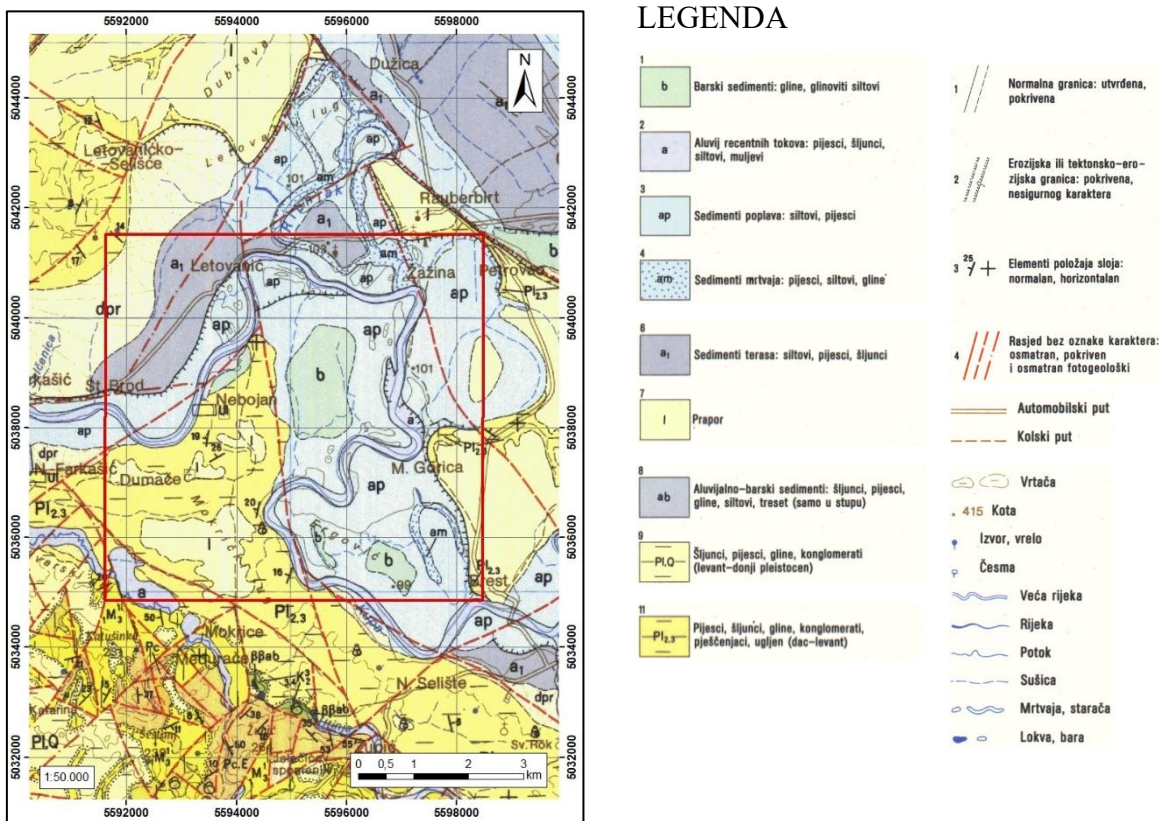
Krajem svibnja 2023. godine dogodile su se poplave na području Karlovačke i Sisačko-moslavačke županije. Poplavom su zahvaćeni Karlovac, Petrinja, Sisak i Hrvatska Kostajnica. Zbog velikih količina oborina došlo je do naglog porasta vodostaja rijeke Kupe, Une, Korane i Save te drugih manjih rijeka i potoka na tom području. Zbog porasta vodostaja rijeke su se izlile iz svojih korita i poplavile su naselja, poljoprivredna zemljišta i prometnice. Poplave su uzrokovale velike materijalne štete. Oštećeni su brojni stambeni i infrastrukturni objekti, poljoprivredna zemljišta i lokalna gospodarstva. Zbog brze reakcije službi spašavanja nije bilo ljudskih žrtava te je stanovništvo na vrijeme evakuirano i privremeno smješteno na sigurna mjesta. Posljedice poplave na području Karlovca prikazane su na slici 2.1.



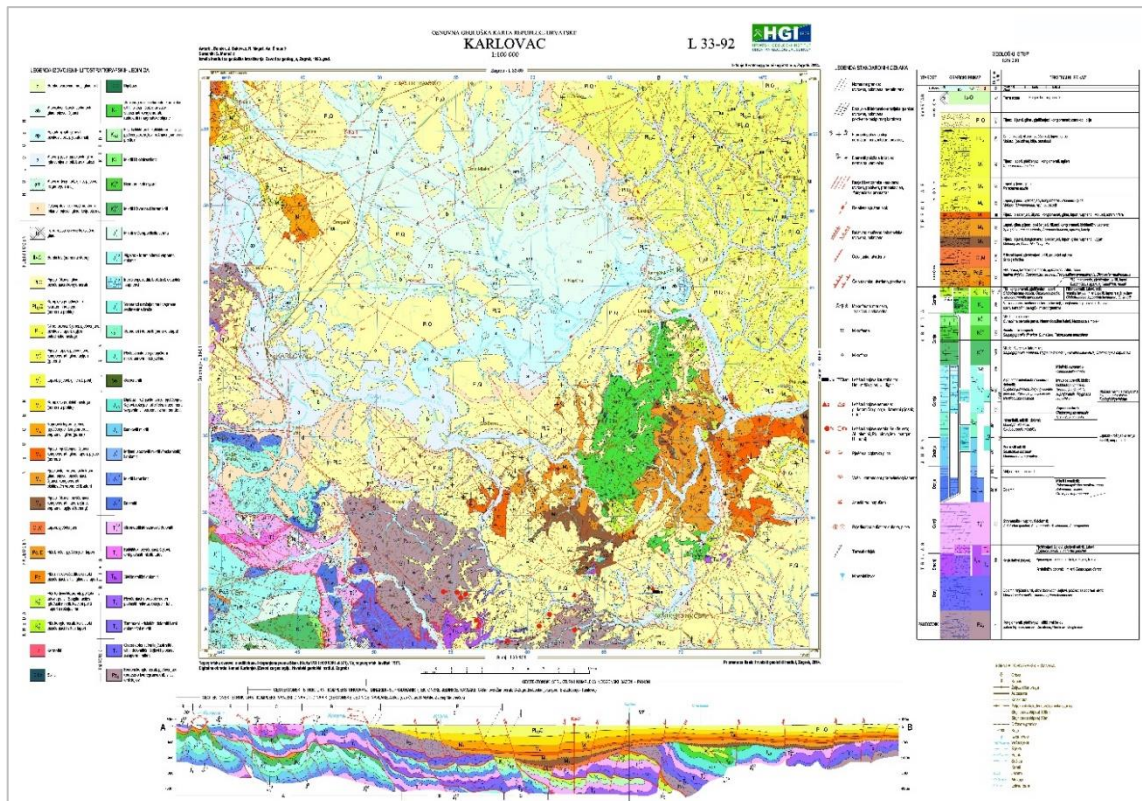
Slika 2.1. Poplava na području Karlovca (Telegram, 2023)

## 2.1. Geologija

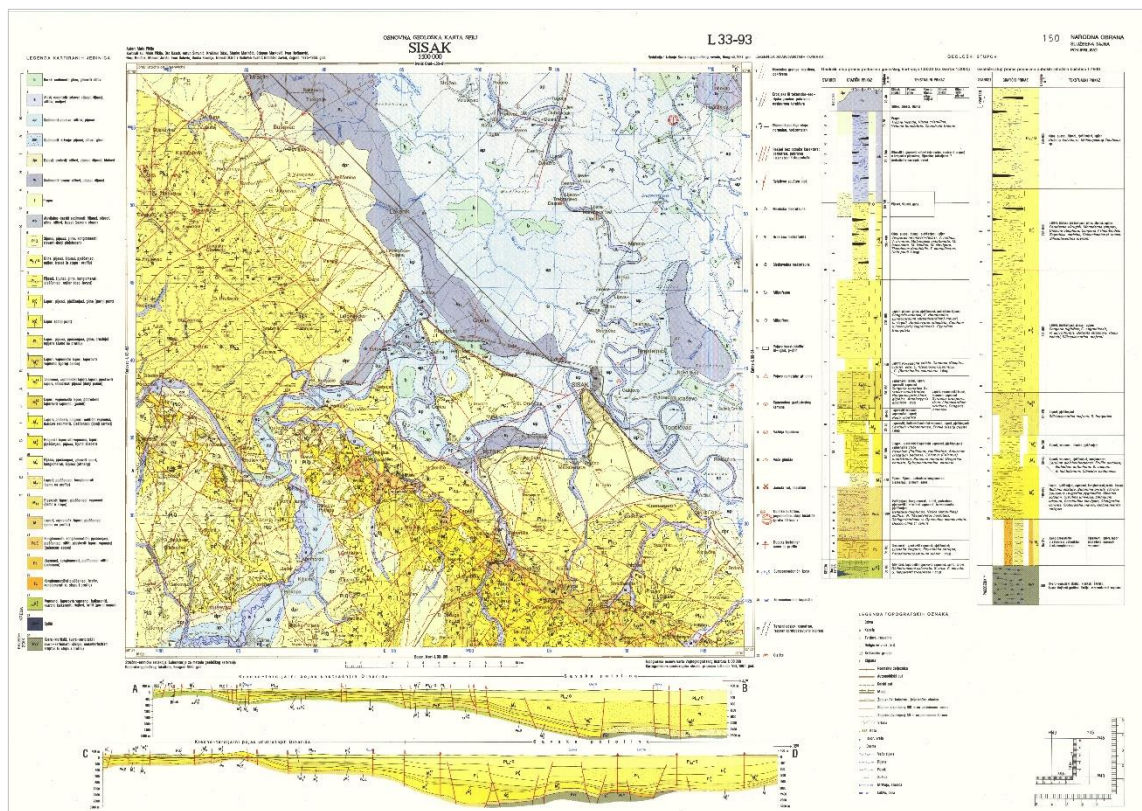
Uvidom u Osnovne Geološke karte (OGK) mjerila 1:100.000 lista Karlovac i Sisak vidljivo je da je na potezu od Karlovca do Siska razvijena geološka struktura oko rijeke Kupe (slika 2.3. i 2.4.). Meandriranjem rijeke Kupe materijal je erodiran i nošen dalje, a zbog tog mehanizma pretežno se talože sedimenti kao što su šljunci, pijesci i gline. Sadržaj gline objašnjava posljedicu plavljenja i zadržavanja vode na tom području. Uvećani prikaz područja istraživanja prikazan je na slici 2.2. Proučavanjem geološke strukture područja istraživanja vidljivo je da oko rijeke Kupe prevladavaju aluvijalni sedimenti odnosno šljunci, pijesci, siltovi i gline (svijetlo do tamno plava boja), a u središtu istraživanog područja postoje i barski sedimenti kao što su gline i glinoviti siltovi (svijetlo zelena boja). Mjestimično je teren građen i od prapora i konglomerata zajedno sa šljuncima, pijescima i glinama (svijetlo do tamno žuta boja). Sve litološko-stratigrafske jedinice su kvartarne starosti, dok je jedinica  $Pl_{2,3}$  tercijarne starosti.



Slika 2.2. Isječak Osnovne geološke karte (OGK), list Sisak (M. Pikija i dr., 1987). Originalno mjerilo OGK je 1:100.000.



Slika 2.3. Osnovna geološka karta, list Karlovac (J. Bukovac, N. Magaš, A. Šimunić i dr., 2014)



Slika 2.4. Osnovna geološka karta, list Sisak (M. Pikija i dr., 1987).

## 2.2. Klimatski uvjeti

Prema podacima DHMZ-a vodostaji rijeka su 17. svibnja 2023. godine ostali visoki zbog kontinuiranih oborina na području gotovo cijele zemlje. Zbog zasićenosti tla i novih oborina, vodostaji rijeka poput Kupe i Korane u Karlovcu te Une u Hrvatskoj Kostajnici dosegli su razine koje su zahtijevale izvanredno stanje obrane od poplava (DHMZ-a, 2024). Zabilježeni su maksimalni vodostaji i protoci na rijekama Korana, Kupa i Una s bliskih hidroloških postaja, a prikazani su u tablici 2.1.

**Tablica 2.1.** Maksimalni vodostaji i protoci na hidrološkim postajama rijeke Korane, Kupe i Une (DHMZ-b, 2024)

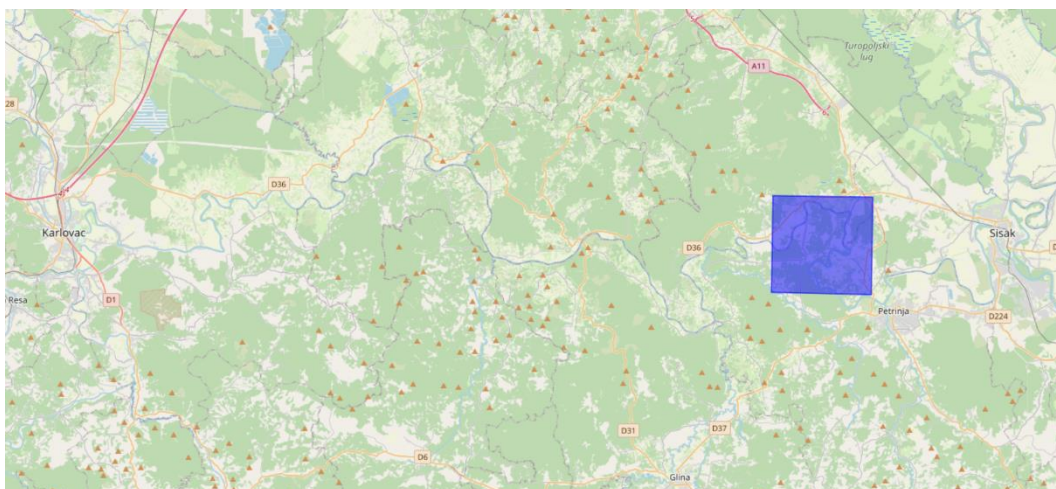
Hidrološka postaja, šifra	Vodostaj (cm) max.	Protok (m <sup>3</sup> /s) max.
Korana		
Velemerić, 4073	524 cm 17.5.2023.	408,8 m <sup>3</sup> /s 17.5.2023.
Karlovac, 4026	844 cm 18.5.2023.	-
Kupa		
Brodarci, 4003	461 cm 17.5.2023.	828,0 m <sup>3</sup> /s 17.5.2023.
Karlovac, 4111	-	833,0 m <sup>3</sup> /s 18.5.2023.
Rečica II, 4056	862 cm 18.5.2023.	1172,0 m <sup>3</sup> /s 18.5.2023.
Jamnička kiselica, 4107	750 cm 19.5.2023.	1260,0 m <sup>3</sup> /s 19.5.2023.
Šišinec, 4065	1056 cm 19.5.2023.	1035,0 m <sup>3</sup> /s 19.5.2023.
Farkašić, 4010	976 cm 19.5.2023.	1311,0 m <sup>3</sup> /s 19.5.2023.
Una		
Kostajnica, 3217	490 cm 17.5.2023.	1813,0 m <sup>3</sup> /s 17.5.2023.



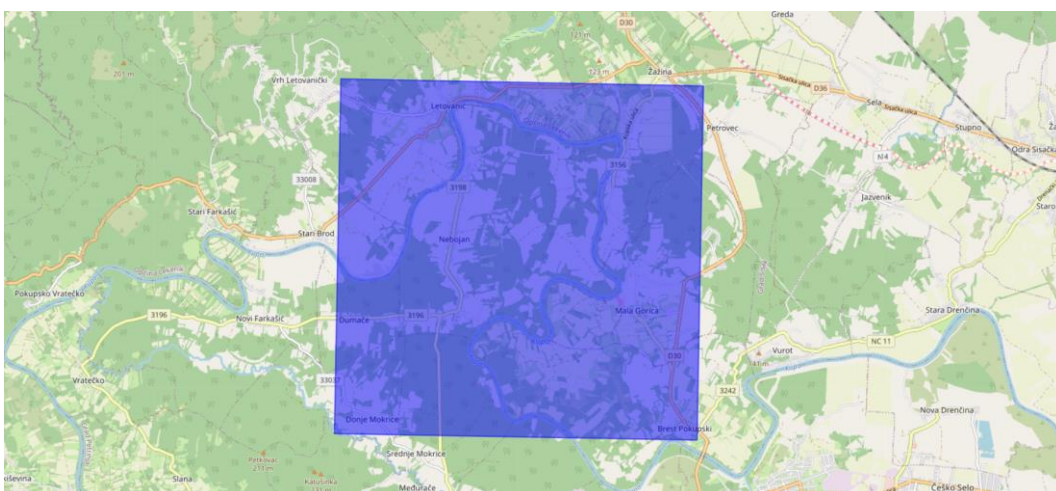
### 3. PODRUČJE ISTRAŽIVANJA

Topografska karta na području Karlovačke i Sisačko-moslavačke županije prikazana je na slici 3.1. Na slici se jasno vidi rijeka Kupa od Karlovca na zapadu do Petrinje i Siska na istoku, a plavi kvadrat označava područje istraživanja.

Budući da su se poplavljena područja protezala duž rijeka čime su poplavljena područja zauzimala velike površine bilo je potrebno odabrati manju površinu odnosno manje područje istraživanja kako bi metoda strojnog učenja bila učinkovita i izvediva. Tako je u svrhu ovog diplomskog rada za područje istraživanja izabran mali dio korita rijeke Kupe čineći jedan kvadrat površine otprilike 30 km<sup>2</sup>. Na slici 3.2. je uvećani prikaz područja istraživanja. Područje istraživanja definirano je kao poligon u ArcMap-u, a zatim je uvezeno u R gdje je prikazano na web karti koristeći paket *Leaflet*.



Slika 3.1. Rijeka Kupa od Karlovca do Siska i označeno područje istraživanja plavim kvadratom



Slika 3.2. Područje istraživanja

## 4. TEORIJSKE OSNOVE

Ovim poglavljem opisuju se teorijske osnove ovog diplomskog rada. Poglavlje je predviđeno kao teoretski vodič za lakše razumijevanje svih navedenih termina.

### 4.1. Copernicus (ESA)

Copernicus je program Europske unije za promatranje Zemlje (Copernicus-1, 2024). Pruža točne i pravovremene informacije za poboljšanje upravljanja okolišem, razumijevanje i ublažavanje posljedica klimatskih promjena i osiguranje civilne sigurnosti. Pružene informacijske usluge su besplatne i lako dostupne korisnicima. Pokriva širok raspon primjena, između ostalog: klimatske promjene, praćenje zemljišta, temperatura, šumskih požara, prirodnih katastrofa (npr. erupcija vulkana, potresi), onečišćenje zraka itd. (Copernicus-2, 2024).

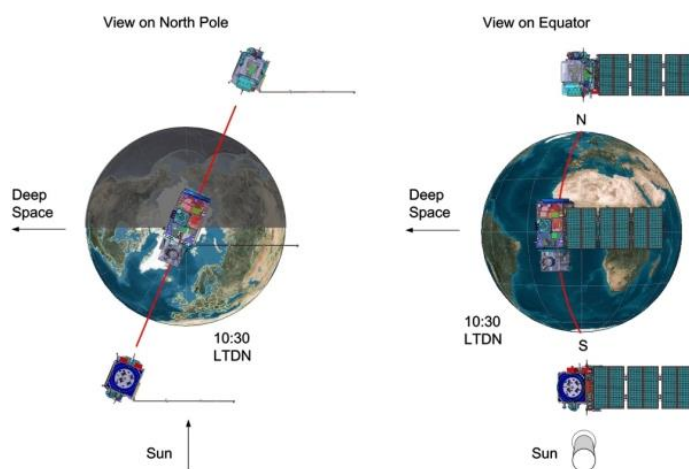
Europska svemirska agencija (ESA - European Space Agency) je razvila skup satelita Sentinel kako bi se zadovoljile operativne potrebe programa Copernicus, Sentinel pruža jedinstven skup promatranja, počevši od radarskih slika za sve vremenske prilike, danju i noću (Sentinel-1), zatim optičke slike visoke rezolucije za kopnene usluge (Sentinel-2) te usluge relevantne za ocean i kopno (Sentinel-3) (Copernicus-2, 2024). Slika 4.1. prikazuje sve misije satelita Sentinel koje su već lansirane i koje se planiraju lansirati u budućnosti.



Slika 4.1. Misije promatranja Zemlje koje je razvila ESA (Copernicus-2, 2024).

### 4.1.1. Sentinel-2

Za ovu svrhu istraživanja koriste se multispektralne satelitske snimke misije Sentinel 2A koji je lansiran 23. lipnja 2015. godine. Misija se temelji na konstelaciji dva identična satelita (Sentinel-2A i Sentinel-2B) u istoj orbiti udaljenih  $180^\circ$  za optimalnu pokrivenost i isporuku podataka (slika 4.2.). Zajedno pokrivaju sve kopnene površine Zemlje, velike otoke, kopnene i obalne vode svakih pet dana na ekvatoru. Sentinel-2 pruža slike visoke razlučivosti u vidljivim i infracrvenim valnim duljinama za praćenje vegetacije, tla i vodenih površina, unutarnjih vodenih puteva i obalnih područja. (Sentinel-a, 2024).



Slika 4.2. Raspored dvaju satelita Sentinel-2 (Sentinel-a, 2024).

Misija uglavnom pruža najučinkovitije informacije za poljoprivrednu i šume. Sentinel-2 se može koristiti za mapiranje promjena u pokrovu zemljišta i za praćenje svjetskih šuma. Također pruža informacije o onečišćenju u jezerima i obalnim vodama. Satelitske snimke poplava, vulkanskih erupcija i klizišta pridonose kartiranju katastrofa i pomažu naporima humanitarne pomoći. Kombinacija visoke razlučivosti (10 m, 20 m i 60 m), novih spektralnih mogućnosti, širine orbitalnog pojasa od 290 km i prolaska satelita svakih 5 dana iznad istog područja pruža kvalitetne i informativne satelitske snimke Zemljine površine (Sentinel-b, 2024).

Sentinel-2 ima multispektralnu kameru (MSI - MultiSpectral Instrument) visoke razlučivosti širokog opsega s 13 spektralnih pojasa odnosno kanala (bands) koji su prikazani u tablici 4.1. s pripadajućom valnom duljinom, rezolucijom i opisom (Sentinel-a, 2024).

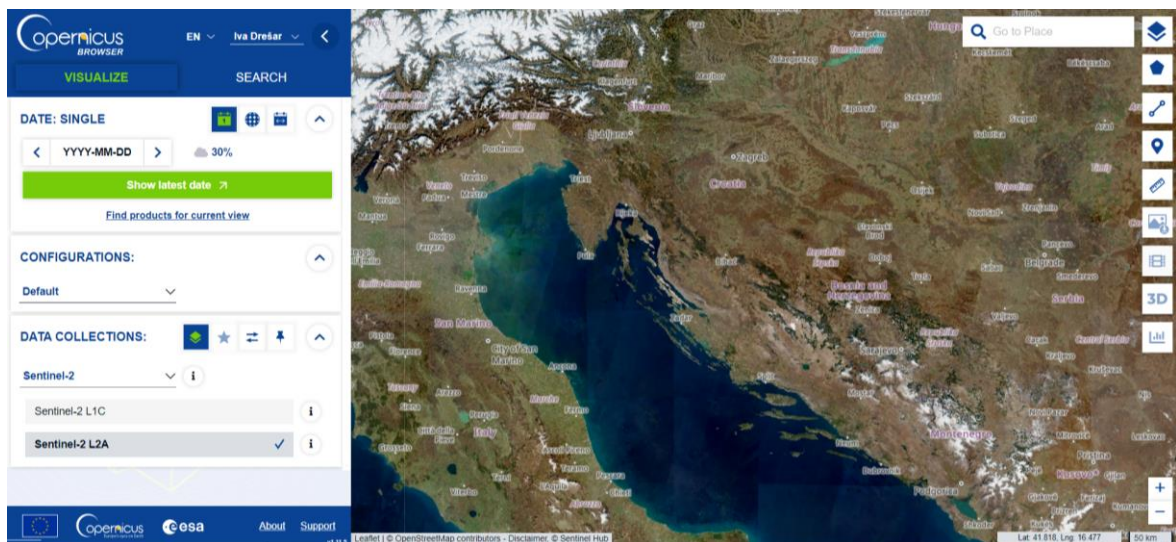
**Tablica 4.1.** Sentinel-2 kanali (Sentinel-a, 2024).

Kanal	Rezolucija (m)	Valna duljina (nm)	Opis	
B1	60	443	Ultra plava (obalna aerosol)	
B2	10	490	Plava (Blue)	RGB kanali
B3	10	560	Zelena (Green)	
B4	10	665	Crvena (Red)	
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	945	Kratkovalni infracrveni (SWIR)	
B10	60	1375	Kratkovalni infracrveni (SWIR)	
B11	20	1610	Kratkovalni infracrveni (SWIR)	
B12	20	2190	Kratkovalni infracrveni (SWIR)	

Prostorna razlučivost Sentinela-2 ovisi o određenom spektralnom pojasu, pa su tako RGB kanali i kanal B8 rezolucije 10 m, dok kanali B5, B6, B7, B8a, B11 i B12 imaju rezoluciju od 20 m. Najmanju prostornu rezoluciju od 60 m imaju kanali B1, B9 i B10 (Sentinel-a, 2024).

### 4.1.2. Geoportal Copernicus

Geoportal Copernicus (Copernicus Data Space Ecosystem Browser) služi kao centralno središte za pristup, istraživanje i korištenje podataka dobivenih promatranjem Zemlje i okoliša, a pružaju ih konstelacije satelita Copernicus Sentinel. Korisnici mogu vizualizirati, uspoređivati, analizirati i preuzimati sve podatke za razne primjene, od praćenja okoliša i upravljanja katastrofama do urbanog planiranja i poljoprivrede. (Copernicus Browser-1, 2024). Na slici 4.3. prikazano je sučelje geoportala Copernicus.



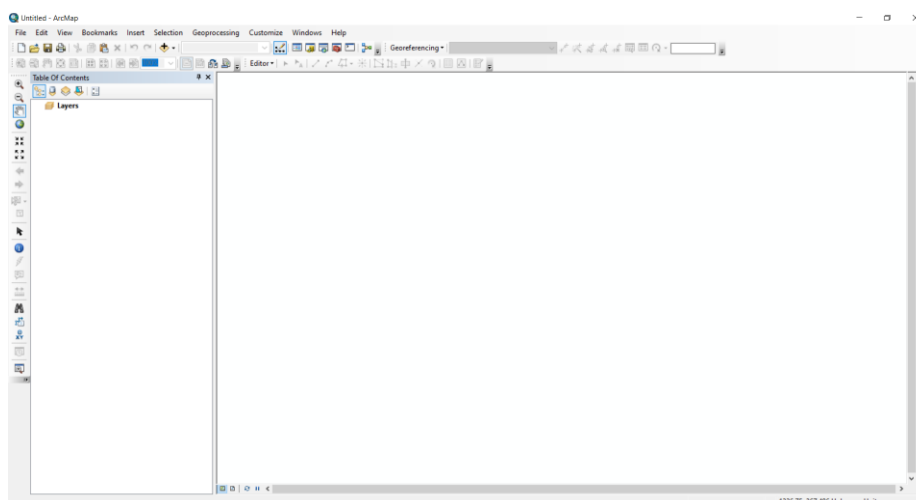
Slika 4.3. Sučelje Copernicus Browser-a (Copernicus Browser-2, 2024).

## 4.2. ArcGIS (GIS) – ESRI

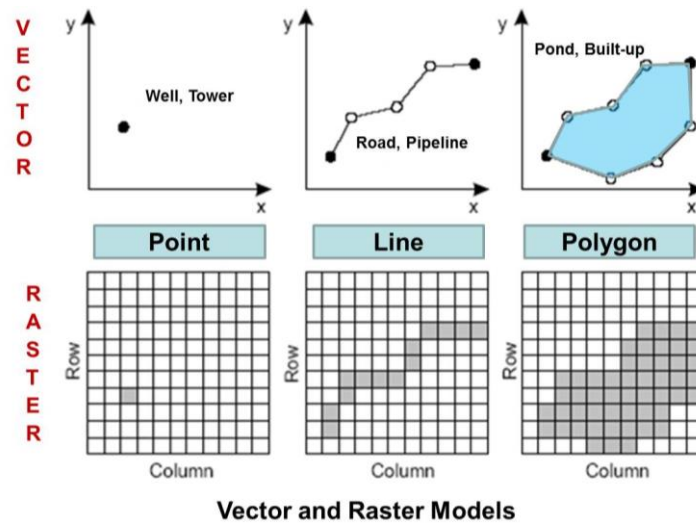
ArcGIS je sveobuhvatna geoprostorna platforma i vodeća je tehnologija geografskog informacijskog sustava (GIS), a kreiran je od strane ESRI-ja (ESRI-1, 2024). GIS je tehnologija koja se koristi za stvaranje, upravljanje, analizu i mapiranje svih vrsta podataka. Povezuje prostorne podatke s kartografskim prikazom. Pruža temelj za mapiranje i analizu, a koristi se u znanosti i gotovo svakoj industriji. Pomaže korisnicima razumjeti obrasce, odnose i geografski kontekst. Prednosti uključuju poboljšanu komunikaciju, učinkovitost, upravljanje i donošenje odluka (ESRI-2, 2024).

### 4.3.1. ArcMap

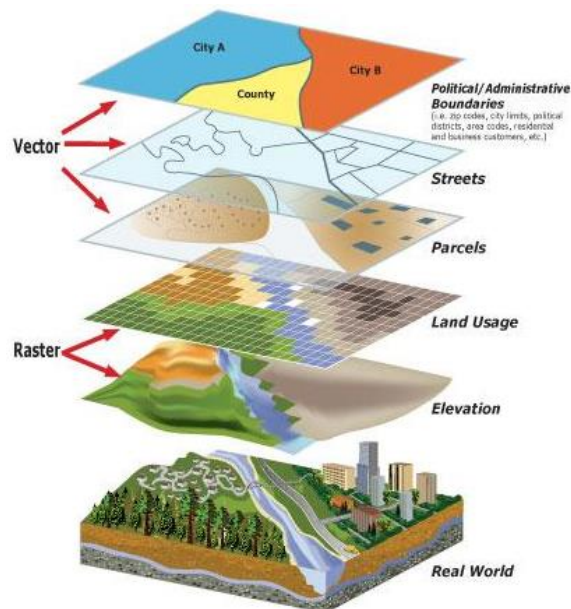
ArcMap je jedna od središnjih aplikacija koje se koriste u ArcGIS Desktopu. ArcMap je mjesto gdje prikazujete i istražujete GIS skupove podataka za određeno područje istraživanja, dodjeljujete simbole i kreirate izgled karte za ispis ili objavljivanje. Također se koristi za stvaranje i uređivanje skupova podataka. ArcMap predstavlja geografske informacije kao skup slojeva i drugih elemenata na karti (ArcGIS Desktop, 2024). Sučelje programa ArcMap prikazano je na slici 4.4. Postoje dvije osnovne grupe podataka koji se koriste u GIS-u, a to su rasterski i vektorski podaci (slika 4.5.). Vektorski podaci za oblikovanje koriste točke, linije ili poligone, dok se rasterski podaci sastoji od mreže ćelija ili piksela, gdje svaka ćelija pohranjuje vrijednost (Medium-a, 2023). Na taj način mogu se u ArcMap uvrštavati vektorskih i rasterskih podataka različitog sadržaja u obliku slojeva (layers) ovisno o vrsti istraživanja, a na slici 4.6. vidimo jedan od primjera takvih vrsta podataka različitog sadržaja.



Slika 4.4. Sučelje programa ArcMap



Slika 4.5. Razlika između rasterskih i vektorskih podataka (GIS, 2024)



Slika 4.6. Vektorski i rasterski podaci različitog sadržaja u obliku slojeva (City of Newberg, 2024)

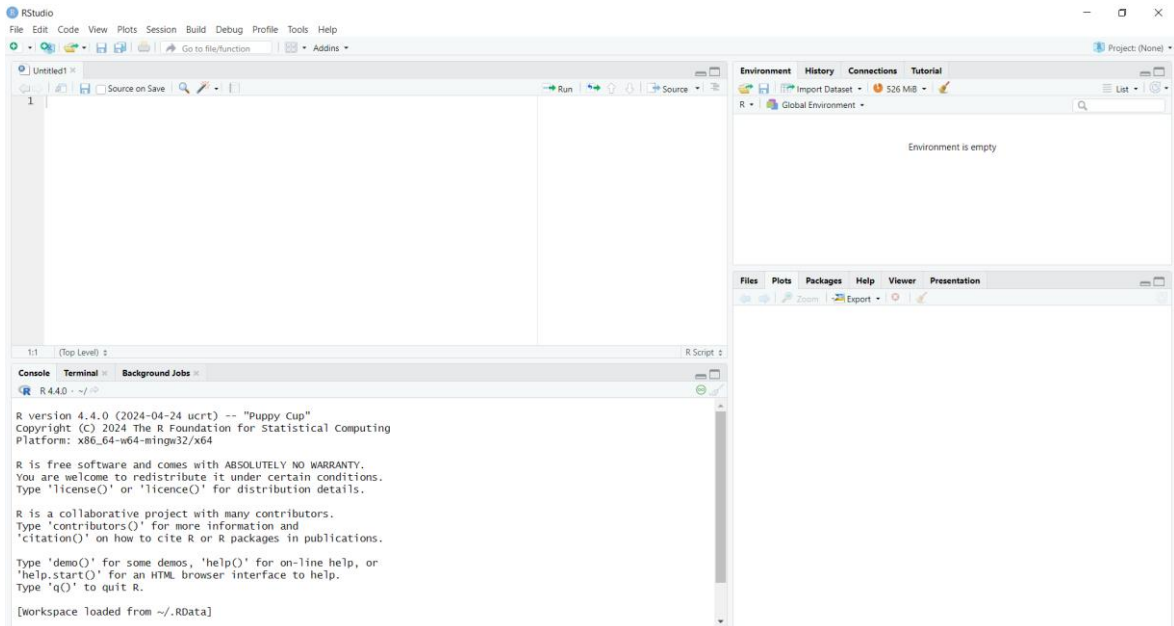
### 4.3. Programski jezik R

R je besplatni softver otvorenog koda i programski jezik koji se koristi za analizu statističkih podataka, a temelji se na starijem S programskom jeziku. R je praktičan i koristan softver zbog toga što je besplatan, dostupan za Windows, Macintosh i Linux, grafički kvalitetan i zbog dostupnosti naprednih statističkih metoda i algoritama objavljenih kao paketi kreirani od strane korisnika. RStudio je integrirano razvojno okruženje (IDE - integrated development environment) za R. IDE pruža sveobuhvatno okruženje za programiranje, izvršavanje i otklanjanje pogrešaka koda u istom softveru. RStudio je jedan od najpopularnijih IDE za R programiranje, a njegova zajednica proizvela je desetke tisuća paketa, s opsegom od strojnog učenja do analize podataka transkriptoma. Poboljšava iskustvo R programiranja jer je sučelje organizirano tako da korisnik može jasno vidjeti u isto vrijeme grafikone, podatkovne tablice, R kod i izlaz (slika 4.7.). Prednost R Studio-a je to što se datoteke mogu učitati bez da se ispisuje kod (Kent State, 2024).

#### 4.2.1. Sučelje R Studio

Sučelje R Studio-a prikazano je na slici 4.7. te se dijeli na četiri prozora. Gornji lijevi prozor predstavlja R skriptu (*R Script*), dok donji lijevi prozor predstavlja konzolu (*Console*). Naredbe se mogu pisati ili izravno u konzoli ili u R skripti koje će se zatim izvršiti u konzoli. Preferirani način izvršavanja naredbi je korištenje R skripte, jer se može spremi i koristiti više puta ili modificirati za izvršavanje u budućnosti. U gornjem desnom prozoru vidljive su kartice *Environment*, *History*, *Connections* i *Tutorial*. Najbitnija je kartica *Environment* koja sadrži sve varijable koje se stvaraju u R-u. U donjem desnom prozoru vidljive su kartice *Files*, *Plots*, *Packages*, *Help*, *Viewer* i *Presentation*. Pomoću kartice *Files* vidljive su sve datoteke u trenutnom radnom direktoriju. Na kartici *Plots* vidljivi su svi iscertani grafikoni i dijagrami, dok se pomoću kartice *Packages* mogu izravno instalirati paketi ili vidjeti svi dostupni paketi. Pomoću kartice *Help* može se provjeriti definiciju i način korištenja ugrađene funkcije. Kartica *Viewer* u RStudio prikazuje web sadržaj generiran unutar R-a, poput HTML izlaza ili vizualizacije stvorene određenim paketima kao što su *shiny*, *leaflet*, *plotly* ili *htmlwidgets*. Kada se pokrene kod koji generira interaktivne vizualizacije ili web-sadržaj, izlaz se prikazuje u ovoj kartici, pružajući interakciju i pregled rezultata izravno unutar R Studio-a. Posebno je korisna za pregled izlaza bez potrebe otvaranja zasebnog web preglednika (Patwal, M., 2020). U ovom diplomskom radu koristi se vizualizacija pomoću paketa *leaflet*.





Slika 4.7. Sučelje R Studio-a

### 4.2.2. R paketi

R paketi su zbirke funkcija, podataka i kompiliranog koda koji proširuju funkcionalnost R-a. Paketi omogućuju korisnicima izvođenje širokog spektra zadataka u R, od manipulacije podacima i vizualizacije do strojnog učenja i specijalizirane statističke analize (R package Wikipedia, 2024). Za instalaciju paketa koristi se naredba *install.packages()*, a upisom naziva paketa u zagrade pokreće se instalacija paketa. Nakon što se paket instalira, njegov sadržaj se može učiniti dostupnim za korištenje pokretanjem naredbe *library()*.

Paketi su podijeljeni u 5 kategorija: prostorna analiza, manipulacija podacima, vizualizacija podataka, strojno učenje i grafika. Prostorna analiza (*Spatial Analysis*) uključuje manipulaciju i analizu prostornih podataka, gdje su podaci povezani s položajem na Zemljinoj površini. U R-u nekoliko paketa je dizajnirano za rad s različitim vrstama prostornih podataka, bilo da se radi o vektorskim (točke, linije, poligoni) ili rasterskim (mreže, pikseli) podacima. Paketi pružaju alate za čitanje, obradu, analizu i vizualizaciju prostornih podataka. Manipulacija podacima (*Data Manipulation*) je ključni element analize podataka, a odnosi se na proces prilagodbe podataka kako bi bili organizirani i lakši za čitanje. Uključuje čišćenje, preoblikovanje i pripremu podataka za analizu. Paketi služe za pojednostavljenje tih procesa. Paketi za vizualizaciju podataka (*Data Visualization*) pružaju alate za stvaranje statičkih i interaktivnih grafikona, pomažući otkriti obrasce, trendove i odnose unutar podataka. Strojno učenje (*Machine Learning*) koristi algoritme za izgradnju modela koji mogu donositi predviđanja ili odluke na temelju podataka, a paketi služe za implementaciju, treniranje, evaluaciju i vizualizaciju modela strojnog učenja u različitim područjima. Kategorija grafika (*Graphics*) sadrži pakete koji podržavaju grafičke operacije u R-u, uključujući stvaranje i prilagodbu grafova te upravljanje uređajima (poput ekrana ili datoteka) gdje se ti grafikoni prikazuju (Moraga, 2023; Wickham i Grolemund, 2017).

U ovom diplomskom radu korišteni su paketi navedeni u tablici 4.2. te su zasebno opisani odnosno opisana je njihova svrha u R programskom jeziku.

**Tablica 4.2.** Korišteni paketi u R programskom jeziku (posit™ Support, 2024; RDocumentation, 2024)

Paketi	Opis
<b><i>Prostorna analiza (Spatial Analysis)</i></b>	
<b>terra</b>	Paket za analizu prostornih podataka s vektorskim (točke, linije, poligoni) i rasterskim ( <i>grid</i> ) podacima.
<b>sp</b>	Paket za prostorne podatke; klase dokumentiraju gdje se nalaze informacije o prostornoj lokaciji, za 2D ili 3D podatke; temelj za prostornu analizu.
<b>sf</b>	Pojednostavljuje rad s vektorskim podacima koristeći standarde jednostavnih značajki; standardizirani način za kodiranje prostornih vektorskih podataka.
<b>raster</b>	Čitanje, pisanje, manipuliranje, analiza i modeliranje prostornih podataka. Rad s rasterskim podacima, kao što su satelitske slike i dr.
<b>leaflet</b>	Omogućuje stvaranje interaktivnih web karata izravno iz R-a.
<b>mapview</b>	Brza i praktična izrada interaktivnih karata vizualizacijom prostornih podataka sa ili bez pozadinskih karata.
<b>leafsync</b>	Sinkronizira više <i>leaflet</i> karata za komparativnu analizu. Kada je sinkronizacija omogućena, sve karte reagiraju na radnje mišem na jednoj karti. Omogućuje usporedbu različitih atributa istih geometrija.
<b><i>Manipulacija podacima (Data Manipulation)</i></b>	
<b>dplyr</b>	Brz i dosljedan alat za rad s objektima, u memoriji i izvan memorije. Bitan za manipulacije podacima poput filtriranja redaka, odabira stupaca i sažimanja podataka.
<b>tidyr</b>	Pojednostavljuje proces sređivanja podataka, čineći ih lakšima za rad u analizama.
<b>data.table</b>	Pruža visokoučinkovitu manipulaciju podacima, posebno za velike skupove podataka. Nudi prirodnu i fleksibilnu sintaksu za brži razvoj.
<b><i>Vizualizacija podataka (Data Visualization)</i></b>	
<b>ggplot2</b>	Sustav za 'deklarativno' stvaranje grafike tj. za stvaranje širokog spektra statičkih vizualizacija, temeljen na "Gramatici grafike".
<b>plotly</b>	Proširuje <i>ggplot2</i> za stvaranje interaktivnih web vizualizacija.
<b>RColorBrewer</b>	Pruža estetski privlačne palete boja za poboljšanje vizualnog dojma karata i grafova, a dizajnirala ih je Cynthia Brewer.

<b><i>Strojno učenje (Machine Learning)</i></b>	
<b>caret</b>	Sveobuhvatan paket koji pojednostavljuje proces strojnog učenja pružajući funkcije za treniranje, podešavanje i evaluaciju modela.
<b>randomForest</b>	Paket za klasifikaciju i regresiju temeljen primjenom algoritma „ <i>random forest</i> “ korištenjem slučajnih ulaza.
<b>MLmetrics</b>	Nudi široki raspon metrika za evaluaciju modela strojnog učenja
<b>LiblineaR</b>	Jednostavan paket za rješavanje linearne klasifikacije i regresije velikog opsega. Glavne značajke <i>LiblineaR-a</i> uključuju klasifikaciju u više klasa, unakrsnu provjeru valjanosti za odabir modela, procjene vjerojatnosti (samo logistička regresija) ili težine za neuravnotežene podatke. Procjena modela je posebno brza u usporedbi s drugim paketima.
<b>NeuralNetTools</b>	Pružila alate za vizualizaciju i analizu koji pomažu u interpretaciji modela neuronskih mreža. Dostupne su funkcije za iscertavanje, kvantificiranje važnosti varijable, provođenje analize osjetljivosti i dobivanje jednostavnog popisa težina modela.
<b><i>Grafika (Graphics)</i></b>	
<b>grDevices</b>	Osnovni paket u R-u za rad s grafičkim uređajima i podrška za osnovnu i mrežnu grafiku.

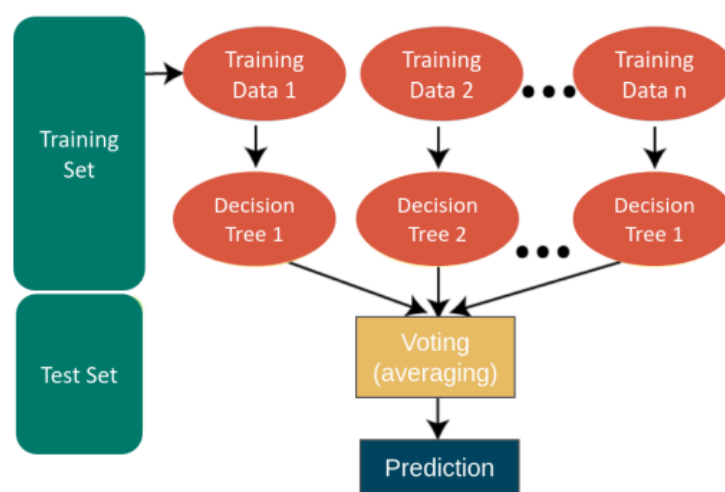
#### 4.2.1. Metode strojnog učenja („Machine learning“)

Strojno učenje (*Machine Learning*) koristi algoritme za izgradnju modela koji mogu donositi predviđanja ili odluke na temelju podataka, a paketi služe za implementaciju, treniranje, evaluaciju i vizualizaciju modela strojnog učenja u različitim područjima (Moraga, 2023; Wickham i Grolemond, 2017). Postoji mnogo algoritama za strojno učenje, ali su za ovaj diplomski rad izabrani algoritmi *Random Forest (RF)*, *Support Vector Machine (SVM)* i *Neural Network (NNET)*.

##### Random Forest (RF)

Nadzirani algoritam strojnog učenja koji se koristi i za klasifikaciju i za regresiju. Algoritam ima veću točnost od drugih algoritama, može odrediti koje su varijable važne u klasifikaciji, učinkovito obrađuje podatke koji nedostaju, šume koje stvori mogu se spremati i koristiti kasnije te može grupirati podatke i pronaći neobične uzorke.

Kako bi algoritam bio funkcionalan, prvo se mora učitati skup podataka koji se zatim dijeli na skup za testiranje (*Test Set*) i skup za treniranje (*Training Set*). Skupovi podataka za treniranje sadrže različite vrijednosti (*Training Data 1, 2, ..., n*). Uz pomoć tih skupova stvara se više modela stabala odlučivanja (*Decision Tree 1, 2, ..., n*). Na temelju rezultata tih modela provodi se glasanje (*Voting*) kako bi se pronašao rezultat s najvećom učestalošću. Skup podataka za testiranje se procjenjuje na temelju dobivenih rezultata kako bi se dobila konačna predikcija (*Prediction*). Opis načina rada algoritma *Random Forest* prikazan je slikom 4.8. (Javapoint, 2021).

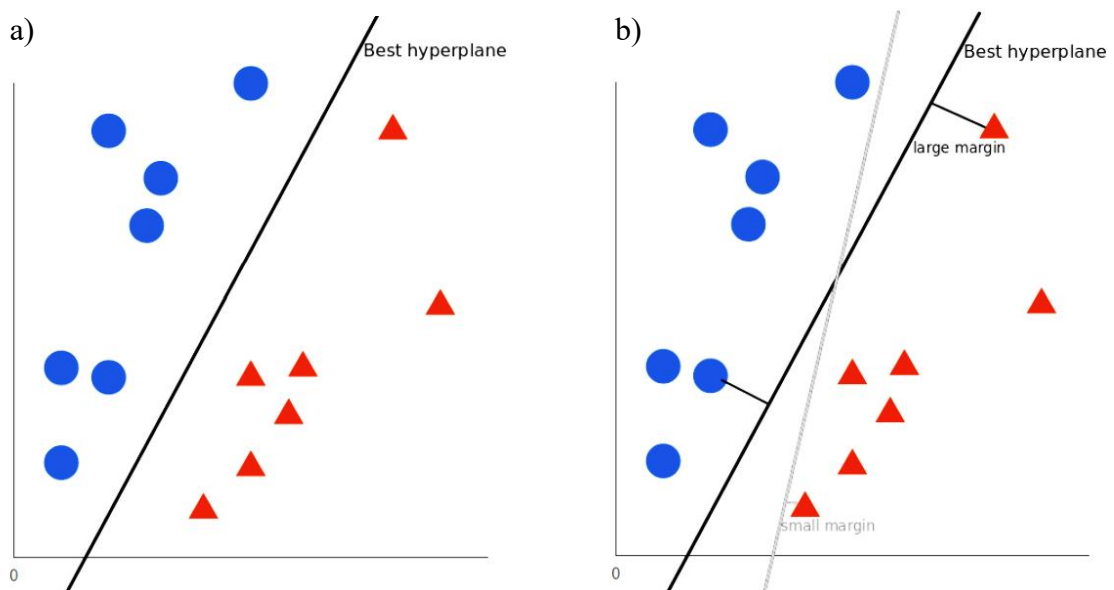


Slika 4.8. Način rada algoritma Random Forest (Javapoint, 2021)

## Support Vector Machine (SVM)

Algoritam analizira podatke koji se koriste za klasifikaciju i regresijsku analizu, ali se uglavnom koristi za klasifikaciju. Cilj *SVM-a* je uzeti skupine opažanja i konstruirati granice kako bi se predvidjelo kojoj skupini tj. klasi pripadaju buduća opažanja na temelju njihovih mjerenja. *SVM* može podnijeti bilo koji broj klasa, kao i promatranja bilo koje dimenzije. Također, može poprimiti gotovo bilo koji oblik (uključujući linearne, radijalne i polinomne oblike) i dovoljno je fleksibilan da se može koristiti u gotovo svim pokušajima klasifikacije koje korisnik odluči poduzeti (UC Business Analytics R Programming Guide, 2024).

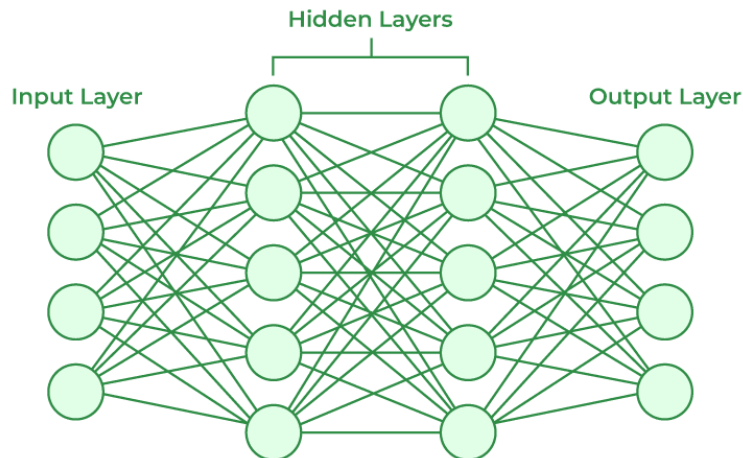
Najjednostavnije objašnjeno, algoritam uzima učitane podatke te vrši klasifikaciju pronalaženjem hiperravnine koja najbolje razdvaja dvije klase (slika 4.9. a)). Hiperravnina predstavlja granicu koju je algoritam stvorio na način da sve što padne na jednu stranu klasificira kao jednu skupinu podataka, a sve što padne na drugu stranu kao drugu skupinu podataka. Postavlja se tamo gdje je udaljenost do najbližeg elementa svake skupine podataka najveća kako je prikazano na slici 4.9. b). Kao i u *RF* metodi strojnog učenja i *SVM* funkcionira tako da se učitani podaci dijele na skup za testiranje i skup za treniranje. Dakle, na temelju skupa za treniranje, algoritam daje optimalnu hiperravninu koja kategorizira nove primjere. Drugim riječima, tijekom procesa obuke modela pronađena je optimalna hiperravnina koja je prvo identificirana u skupu podataka za treniranje, a zatim je potvrđena kao najbolja u skupu podataka za testiranje (Datacamp, 2018).



Slika 4.9. Način rada algoritma Support Vector Machine (Datacamp, 2018)

## Neural Network (NNET)

Neuronska mreža je računalni model inspiriran strukturom i funkcijom ljudskog mozga, a općenito se koristi za rješavanje složenih problema kao što su klasifikacija slika, prepoznavanje govora i generiranje teksta. Sastoji se od međusobno povezanih čvorova, neurona, organiziranih u slojeve. Mreža prima ulazne podatke, obrađuje ih kroz više slojeva neurona te proizvodi izlaz ili predviđanje (slika 4.10.). Osnovna građevna jedinica neuronske mreže je neuron koji predstavlja računsku jedinicu. Neuronska mreža obično se sastoji od tri vrste slojeva: ulaznog (*Input Layer*), skrivenog (*Hidden Layers*) i izlaznog sloja (*Output Layer*). Prvi sloj prima neobrađene podatke, obrađuje ih više skrivenih slojeva, a izlazni sloj daje rezultat. Neuronske mreže su fleksibilne i mogu se koristiti i za probleme regresije i klasifikacije. Svaki podatak koji se može učiniti numeričkim može se koristiti u modelu *NNET*, budući da je to matematički model s funkcijama aproksimacije. Prikladne su za modeliranje s nelinearnim podacima s velikim brojem ulaza (npr. slike) jer funkcioniraju tako da problem klasifikacije dijeli na slojevitu mrežu jednostavnijih elemenata. Također, mogu se trenirati s bilo kojim brojem ulaza i slojeva, a najbolje rade s većom količinom podataka (Geeksforgeeks, 2023; Packtpub, 2024).



**Slika 4.10.** Način rada algoritma Neural Network (Geeksforgeeks, 2023)

## 5. PRIPREMA PODATAKA

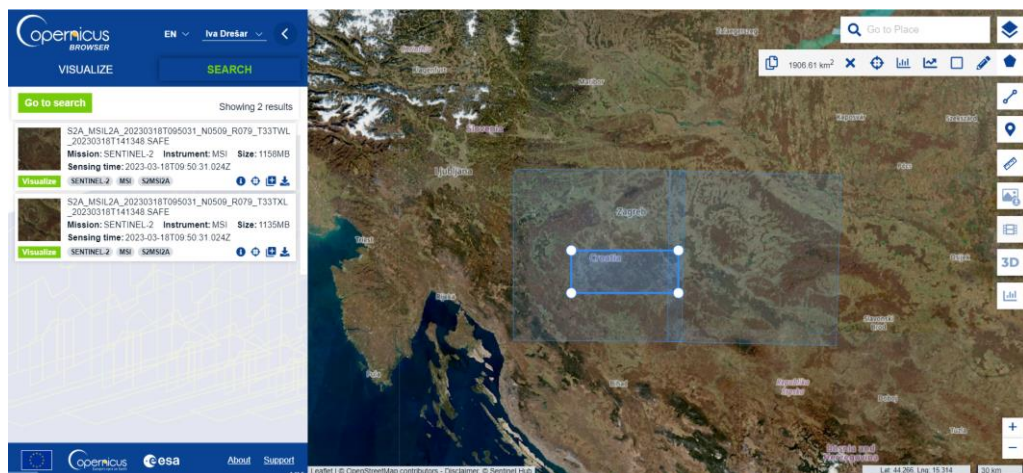
Na stranici geoportal Copernicus preuzete su multispektralne satelitske snimke satelita Sentinel-2 na način da se iscrtava poligon na željeno područje istraživanja. Zatim se odabire izvor podataka (*Data source*), a u ovom slučaju je to Sentinel-2 misije L2A. Nakon toga se odredi vremenski raspon (*Time range*) tako da se dobiju satelitske snimke prije i poslije poplave te maksimalna pokrivenost naoblakom do 10% (*Max. cloud coverage*) (slika 5.1.).

The screenshot displays the Copernicus Browser search interface. At the top, there is a header with the Copernicus logo, language selection (EN), user name (Iva Drešar), and navigation arrows. Below the header, there are buttons for 'VISUALIZE' and 'SEARCH'. The main content area is divided into three sections: 'SEARCH CRITERIA', 'DATA SOURCES', and 'TIME RANGE'. The 'SEARCH CRITERIA' section has a text input field for 'Product name' and a note: 'To apply a location filter, please define an AOI/a POI'. The 'DATA SOURCES' section lists various satellite products with checkboxes and 'Filters' links. 'SENTINEL-2' is selected, and its sub-products 'MSI', 'L1C', and 'L2A' are also selected. A cloud coverage slider is set to 10%. Other products listed include SENTINEL-1, SENTINEL-3, SENTINEL-5P, SENTINEL-6, CCM Optical, CCM DEM, CCM SAR, and GLOBAL-MOSAICS. The 'TIME RANGE' section has 'From' and 'Until' date and time pickers. The 'From' date is 2023-03-01 and the 'Until' date is 2023-03-31. There is also a 'Filter by months' checkbox and a 'Search' button at the bottom. The footer contains logos for the European Union, Copernicus, and ESA, along with 'About' and 'Support' links and the version number 'v1.11.0'.

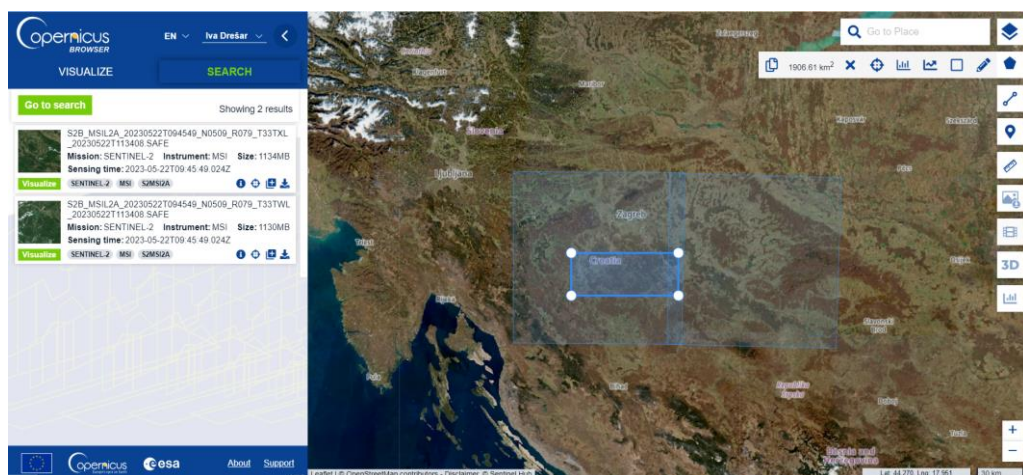
Slika 5.1. Odabir satelitske misije, maksimalne pokrivenosti naoblakom i vremenskog raspona (Copernicus Browser-2, 2024).



Geoportal Copernicus pretražuje multispektralne satelitske snimke tako da odgovaraju traženim zahtjevima. Na taj način dobivaju se satelitske snimke prostorne rezolucije od 10 m, 20 m i 60 m. Satelitske snimke koje predstavljaju istraživano područje prije poplave odabrane su proizvoljno u 3. mjesecu (18.3.2024.) kada sa sigurnošću nije bilo većih oborina koje bi uzrokovale poplave. Najveća poplava bila je oko datuma 17.5.2023., ali u to vrijeme nema satelitskih snimaka s naoblakom manjom od 80%. Prve satelitske snimke koje imaju pokrivenost naoblakom manju od 10% su snimljene na datum 22.7.2024. Bilo bi idealno kada bi postojale snimke oko datuma 17.5. kada su bili najveće poplave, ali zbog velike pokrivenosti oblacima to nije moguće. Od ponuđenih satelitskih snimaka prije i poslije poplave izabrani su prvi ponuđeni podaci jer pokrivaju cijelo i šire područje istraživanja (slika 5.2. i 5.3.).



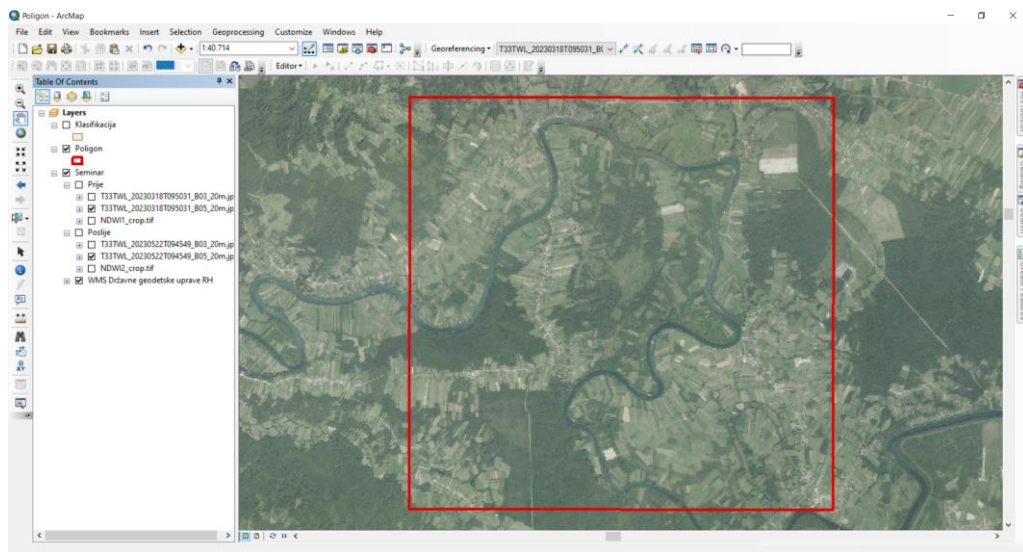
**Slika 5.2.** Multispektralne satelitske snimke prije poplave (18.3.2023.) (Copernicus Browser-2, 2024)



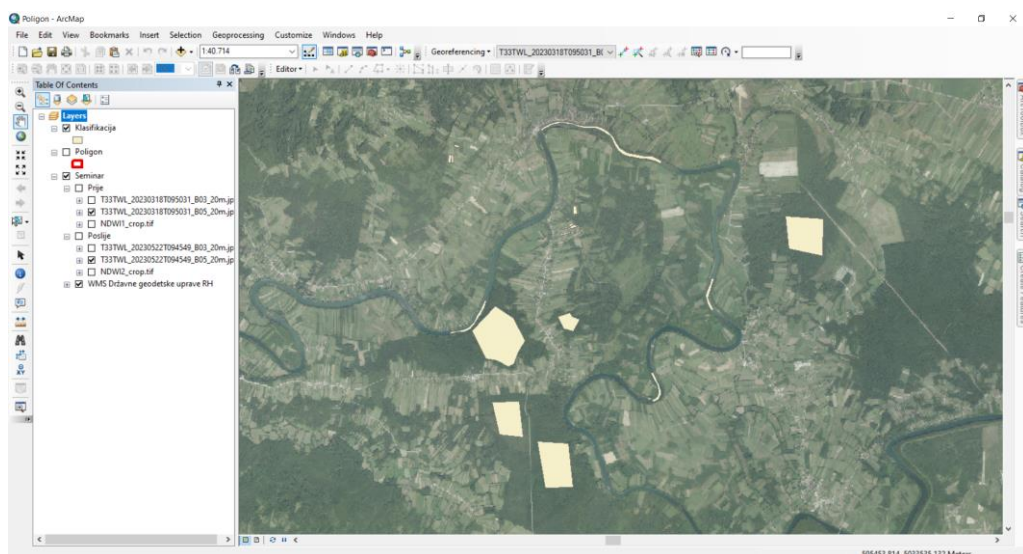
**Slika 5.3.** Multispektralne satelitske snimke poslije poplave (22.5.2023.) (Copernicus Browser-2, 2024)

U poglavlju 4.1.1. je navedeno kako se multispektralna satelitska snimka sastoji od 13 kanala, a u ovom diplomskom radu se koristi 9 kanala: B2, B3, B4, B6, B7, B8, B8a, B11, B12. Koristili su se samo kanali 10 i 20 metarske rezolucije, a kanal B5 koji je 20 metarske rezolucije nije uključen u preuzetim podacima s Copernicus Browser-a.

Korištenjem multispektralnih satelitskih snimaka Sentinel-2 određeno je područje istraživanja u programu ArcMap (crveni kvadrat) (slika 5.4.). Naknadno su dodani poligoni kao ulazni podatak za strojno učenje, a iscrtano je 27 poligona (slika 5.5.). Svakom poligonu je pridodana klasa poljoprivreda, infrastruktura, šuma ili voda, a klase su definirane u atributnoj tablici (slika 5.6.).



Slika 5.4. Područje istraživanja definirano u programu ArcMap



Slika 5.5. Poligoni za strojno učenje definirani u programu ArcMap

Table

Klasifikacija

FID	Shape *	Id	Namjena
0	Polygon	1	Šume
1	Polygon	1	Šume
2	Polygon	1	Šume
3	Polygon	1	Šume
4	Polygon	1	Šume
5	Polygon	2	Pojoprivreda
6	Polygon	2	Pojoprivreda
7	Polygon	2	Pojoprivreda
8	Polygon	2	Pojoprivreda
9	Polygon	2	Pojoprivreda
10	Polygon	2	Pojoprivreda
11	Polygon	2	Pojoprivreda
12	Polygon	2	Pojoprivreda
13	Polygon	3	Voda
14	Polygon	3	Voda
15	Polygon	3	Voda
16	Polygon	3	Voda
17	Polygon	4	Infrastruktura
18	Polygon	4	Infrastruktura
19	Polygon	4	Infrastruktura
20	Polygon	4	Infrastruktura
21	Polygon	4	Infrastruktura
22	Polygon	4	Infrastruktura
23	Polygon	4	Infrastruktura
24	Polygon	4	Infrastruktura
25	Polygon	4	Infrastruktura
26	Polygon	4	Infrastruktura

(0 out of 27 Selected)

Klasifikacija

Slika 5.6. Atributna tablica sa definiranim klasama za svaki poligon

## 6. OBRADA PODATAKA

Treba istaknuti da se u ovom poglavlju detaljno opisuje postupak obrade podataka u programskom jeziku R na satelitskim snimkama prije poplave. Isti postupak primijenjen je i na satelitskim snimkama nakon poplave. Prikazivati će se kod samo za obradu podataka prije poplave, a rezultati, grafovi, dijagrami i slike će se prikazivati za oba slučaja.

U programskom jeziku R najprije se postavlja radni direktorij pomoću naredbi *setwd* i *getwd*, a iz njega se povlače preuzete satelitske snimke te će se također u njemu spremati datoteke. Naknadno su preuzeti paketi pomoću naredbe *install.packages*, a sa naredbom *library* se pozovu paketi. Svi paketi koji su preuzeti su navedeni u poglavlju 4.2.2. sa detaljnim objašnjenjem kojoj kategoriji pripadaju zajedno sa opisom svrhe svakog paketa.

```
setwd("C:/Users/SHU/Desktop/Diplomski/Prije poplave")
getwd()

install.packages("sp")
library(sp)
install.packages("sf")
library(sf)
install.packages("raster")
library(raster)
install.packages("leaflet")
library(leaflet)
install.packages("terra")
library(terra)
```

Sljedeći korak uključuje učitavanje i ponovno uzorkovanje satelitskih snimaka iz 10 metarske rezolucije u 20 metarsku rezoluciju piksela. Odabir kanala, prema kojem će se izvesti ponovno uzorkovanje, nije ključan jer se preuzimaju samo meta podaci tako da kod pretvaranja iz jedne rezolucije u drugu ostane isti koordinatni sustav, dimenzija, broj i pozicija svakog piksela. Ponovno uzorkovanje je izvedeno za kanale B03 i B08 jer nedostaju u folderu R20m sa 20 metarskim kanalima. Novi rasteri spremljeni su u obliku tif datoteke.











```

direktorij <- "C:/Users/SHU/Desktop/Diplomski/Snimke/Prije poplave/"
f1a <- "R10m/T33TWL_20230318T095031_B03_10m.jp2"
f2b <- "R20m/T33TWL_20230318T095031_B02_20m.jp2"
B03_10 <- rast(paste0(direktorij,f1a))
B02_20 <- rast(paste0(direktorij,f2b))
B03_20 <- resample(B03_10, B02_20, method='near') #bilinearna funkcija
f3b <- as.character(gsub("B02", "B03", f2b))
f3b_tif <- as.character(gsub(".jp2", ".tif", f3b))
terra::writeRaster(B03_20, paste0(direktorij,f3b_tif), overwrite = T)

direktorij <- "C:/Users/SHU/Desktop/Diplomski/Snimke/Prije poplave/"
f1a <- "R10m/T33TWL_20230318T095031_B08_10m.jp2"
f2b <- "R20m/T33TWL_20230318T095031_B02_20m.jp2"
B08_10 <- rast(paste0(direktorij,f1a))
B02_20 <- rast(paste0(direktorij,f2b))
B08_20 <- resample(B08_10, B02_20, method='near') #bilinearna funkcija
f3b <- as.character(gsub("B02", "B08", f2b))
f3b_tif <- as.character(gsub(".jp2", ".tif", f3b))
terra::writeRaster(B08_20, paste0(direktorij,f3b_tif), overwrite = T)

```

Slika 6.1. prikazuje sve kanale u folderu R20m koji će se koristiti za terensku klasifikaciju, uključujući nove rastere u tif formatu. Prikazani su kanali satelitskih snimaka prije poplave, a iste takve datoteke su pripremljene i za satelitske snimke poslije poplave.

 T33TWL_20230318T095031_B01_20m	18.3.2023. 16:29	JP2 datoteka	18.757 KB
 T33TWL_20230318T095031_B02_20m	18.3.2023. 16:29	JP2 datoteka	32.903 KB
 T33TWL_20230318T095031_B03_20m	11.7.2024. 9:20	TIF datoteka	56.262 KB
 T33TWL_20230318T095031_B04_20m	18.3.2023. 16:29	JP2 datoteka	32.991 KB
 T33TWL_20230318T095031_B06_20m	18.3.2023. 16:29	JP2 datoteka	33.003 KB
 T33TWL_20230318T095031_B07_20m	18.3.2023. 16:29	JP2 datoteka	32.959 KB
 T33TWL_20230318T095031_B08_20m	11.7.2024. 9:21	TIF datoteka	68.999 KB
 T33TWL_20230318T095031_B8A_20m	18.3.2023. 16:29	JP2 datoteka	33.003 KB
 T33TWL_20230318T095031_B11_20m	18.3.2023. 16:29	JP2 datoteka	32.997 KB
 T33TWL_20230318T095031_B12_20m	18.3.2023. 16:29	JP2 datoteka	32.973 KB

Slika 6.1. Folder R20m sa kanalima za terensku klasifikaciju

Nakon što su kanali pripremljeni u 20 metarskoj rezoluciji učitavaju se svi kanali odnosno rasteri te se slažu u jedan višekanalni rasterski sloj (*Raster Stack object*) koji se zatim sprema kao nova tif datoteka (*KANALI.tif*).

```
R20m_kanali <- c("02_20m.jp2", "03_20m.tif", "04_20m.jp2", "06_20m.jp2",
                "07_20m.jp2", "08_20m.tif", "8A_20m.jp2", "11_20m.jp2",
                "12_20m.jp2")
filenames <- paste0("C:/Users/SHU/Desktop/Diplomski/Snimke/Prije poplave//R20m/
                    T33TWL_20230318T095031_B", R20m_kanali)
s1 <- rast(filenames)
terra::writeRaster(s1, "C:/Users/SHU/Desktop/Diplomski/Prije poplave/KANALI.tif"
                  overwrite = T)
```

U nastavku se učitavaju podaci za strojno učenje, definirani u ArcMap programu, a to su područje istraživanja (*Poligon.shp*) i poligoni za treniranje (*Klasifikacija.shp*) u shp formatu. Uz njih se učitava i prethodno definirani rasterski sloj (*KANALI.tif*).

```
POLIGONI <- st_read("AOI/Klasifikacija.shp") #poligoni za treniranje
sm <- rast("KANALI.tif")
AOI <- st_read("AOI/Poligon.shp") #Area Of Interest - Područje istraživanja
```

Nadalje se rasterski sloj reže prema definiranom području istraživanja naredbom *crop*, a izrezani rasterski sloj se sprema pod novi naziv *KANALIA.tif*. Naredba *rast* učitava novi rasterski sloj i sprema ga pod varijablu *s1*.

```
srtm_2 <- mask(crop(as(sm, "Raster"), AOI), AOI)
writeRaster(srtm_2, filename="KANALIA.tif", overwrite=TRUE)
s1 <- rast("KANALIA.tif")
```

Ovaj kod provodi analizu prostornih podataka korištenjem postupaka filtriranja i pripreme podataka za daljnju obradu. Najprije se izdvajaju jedinstvene vrijednosti/klase unutar atributne tablice shapefile-a (POLIGONI – Klasifikacija.shp) koji sadrži poligone namijenjene za treniranje modela klasifikacije. U ovom slučaju, jedinstvene vrijednosti predstavljaju različite klase zemljišta: poljoprivreda, infrastruktura, šume i vodene površine. Nakon toga, definirano je područje interesa unutar kojeg se podaci filtriraju kako bi se izdvojile samo relevantne prostorne informacije. Daljnja obrada uključuje uklanjanje nepotrebnih podataka kako bi se optimizirala učinkovitost rada.

```

unique(POLIGONI$Namjena)
sm_box_sf <- st_as_sf(st_bbox(s1))
odl_1 <- st_filter(POLIGONI , sm_box_sf)
rm(POLIGONI, sm_box_sf)

odl_1 <- st_filter(odl_1, AOI)
unique(odl_1$Namjena)

names(s1) <- c("B02", "B03", "B04", "B06", "B07", "B08", "B8A", "B11", "B12")|

```

Zatim se prostorni podaci pretvaraju u tablični oblik radi lakše manipulacije. Podaci se iz različitih izvora spajaju u jednu tablicu, s naglaskom na očuvanju relevantnih informacija.

```

d <- terra::extract(s1, odl_1, touches=T)
d <- d %>% drop_na()

odl_1_dt <- tibble(st_drop_geometry(odl_1))
d <- tibble(d)
odl_1_dt$ID1 <- 1:nrow(odl_1_dt)

do <- merge(odl_1_dt, d, by.x = "ID1", by.y = "ID", all.y = T)
names(do)

d01 <- do %>% dplyr::select(-one_of("ID1", "ID"))
d01 <- rename(d01, class = Namjena )
d01 <- d01 %>% relocate(class, .after = last_col())

dt <- d01
dt <- as.data.table(dt)

```

Na taj način stvorena je tablica koja sadrži vrijednosti za svaki piksela iz različitih spektralnih kanala odnosno rastera (satelitskih snimaka) koje su izdvojene na temelju poligona koji predstavljaju klasifikaciju terena: infrastruktura, poljoprivreda, šume i voda. Zbog velike količine podataka prikazan je samo početak i kraj tablice. Tablica 6.1. služi kao ulazni podatak kod algoritama za strojno učenje na kanalima prije poplave, dok je tablica 6.2. za kanale poslije poplave.

**Tablica 6.1.** Prikaz vrijednosti piksela iz različitih kanala s pridodanom klasom – prije poplave

	<b>B02</b>	<b>B03</b>	<b>B04</b>	<b>B06</b>	<b>B07</b>	<b>B08</b>	<b>B8A</b>	<b>B11</b>	<b>B12</b>	<b>class</b>
<b>1</b>	1321	1478	1695	2357	2591	2654	2885	3519	2694	Šume
<b>2</b>	1406	1590	1828	2317	2516	2907	2812	3515	2718	Šume
<b>3</b>	1341	1472	1724	2339	2547	2649	2834	3519	2700	Šume
<b>4</b>	1288	1407	1654	2310	2482	2670	2802	3335	2570	Šume
<b>5</b>	1245	1348	1527	2245	2446	2527	2773	3242	2469	Šume
...										
<b>4868</b>	1560	1814	1951	2338	2472	2672	2633	3297	2940	Infrastruktura
<b>4869</b>	1368	1683	1640	3048	3272	3088	3452	3197	2536	Infrastruktura
<b>4870</b>	1569	1858	1880	2579	2975	3776	3175	3334	3006	Infrastruktura
<b>4871</b>	1435	1477	1759	2139	2375	2578	2734	3244	3060	Infrastruktura
<b>4872</b>	1569	1858	1880	2579	2975	3776	3175	3334	3006	Infrastruktura



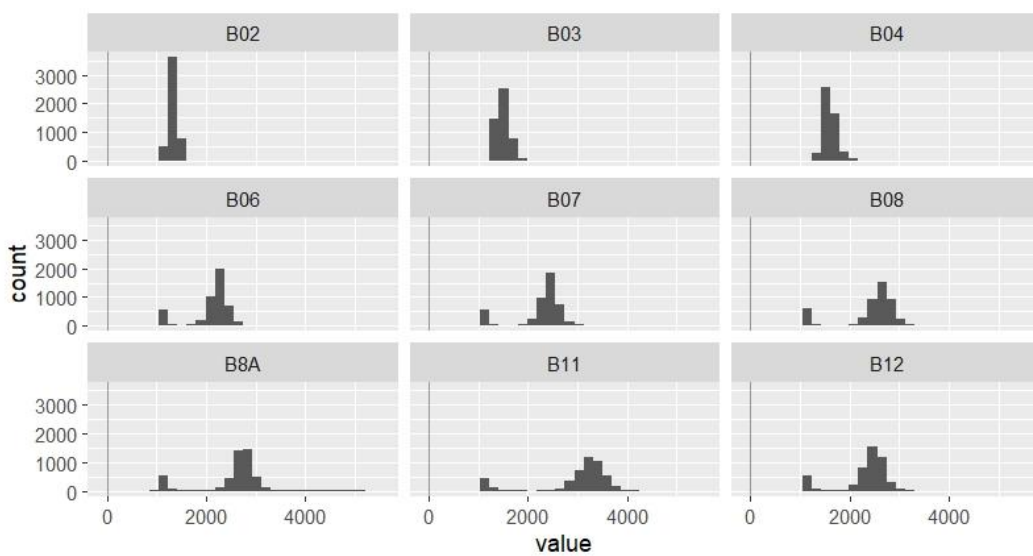
**Tablica 6.2.** Prikaz vrijednosti piksela iz različitih kanala s pridodanom klasom – poslije poplave

	<b>B02</b>	<b>B03</b>	<b>B04</b>	<b>B06</b>	<b>B07</b>	<b>B08</b>	<b>B8A</b>	<b>B11</b>	<b>B12</b>	<b>class</b>
<b>1</b>	1308	1512	1251	3989	4725	4782	4948	2726	1734	Šume
<b>2</b>	1326	1540	1271	3906	4757	4808	4859	2726	1754	Šume
<b>3</b>	1326	1520	1266	3969	4815	4647	5026	2742	1735	Šume
<b>4</b>	1308	1470	1244	3859	4699	4460	4843	2742	1743	Šume
<b>5</b>	1291	1462	1234	3855	4644	4587	4821	2683	1725	Šume
...										
<b>4868</b>	1725	1928	2068	2830	3139	3352	3338	3676	3271	Infrastruktura
<b>4869</b>	1594	1760	1813	3173	3546	3226	3707	3635	2977	Infrastruktura
<b>4870</b>	1716	2060	1979	2786	2867	3293	3258	3798	3481	Infrastruktura
<b>4871</b>	1749	1888	2084	2591	2731	2633	3030	3812	3725	Infrastruktura
<b>4872</b>	1716	2060	1979	2786	2867	3293	3258	3798	3481	Infrastruktura

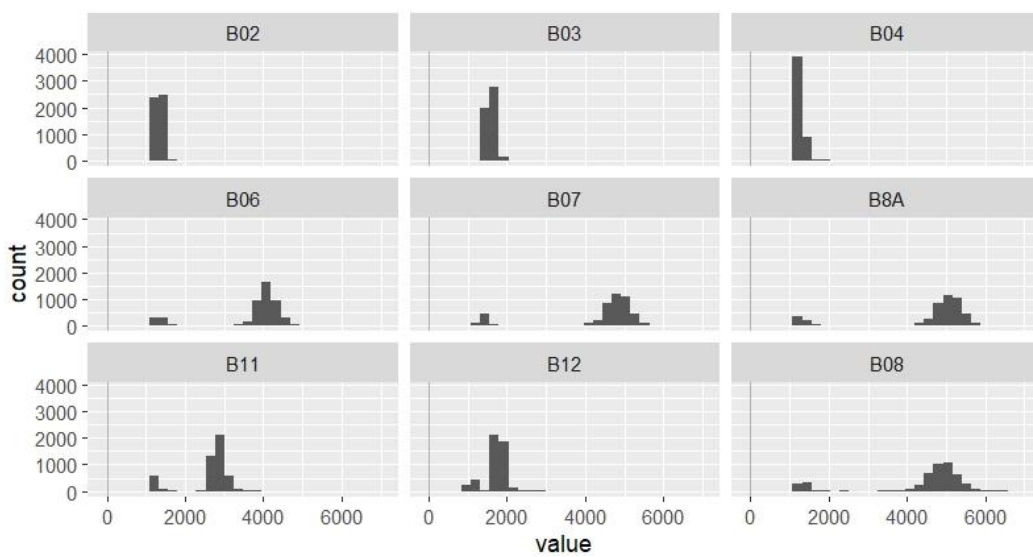
Kombinacijom nekoliko funkcija dobiva se vizualna raspodjela podataka iz tablice 6.1. u obliku histograma za svaki kanal (slika 6.2.). Slika 6.3. prikazuje histograme svakog kanala multispektralnih snimaka poslije poplave za tablicu 6.2.

```
dt %>%
  dplyr::select(-"class") %>%
  melt(measure.vars = names(.)) %>%
  ggplot() +
  geom_histogram(aes(value)) +
  geom_vline(xintercept = 0, color = "gray70") +
  facet_wrap(facets = vars(variable), ncol = 3)

unique(dt$class)
```



Slika 6.2. Histograme za svaki kanal multispektralnih snimaka – prije poplave.



Slika 6.3. Histograme za svaki kanal multispektralnih snimaka – poslije poplave

Korištenjem funkcije *createDataPartition* podaci iz tablice 6.1. se dijele na podatke za treniranje i podatke za testiranje. Za treniranje će biti izdvojeno 70% podataka ( $p=0,7$ ), dok će preostalih 30% biti korišteno za testiranje. Tablice 6.3. i 6.4. prikazuju koliko će elemenata biti za treniranje, a koliko za testiranje te vrijede za satelitske snimke prije i poslije poplave. Također, prije nego što se podaci podijele na treniranje i testiranje potrebno je sa naredbom *set.seed(321)* osigurati da je kod ponovljiv i pouzdan, tj. da svaki put kada se izvrši daje isti rezultat.

```
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 6.3.** Podaci za treniranje – prije i poslije poplave

<i>table(dt_train\$class)</i>			
Infrastruktura	Poljoprivreda	Šume	Voda
26	110	2814	463

**Tablica 6.4.** Podaci za testiranje – prije i poslije poplave

<i>table(dt_test\$class)</i>			
Infrastruktura	Poljoprivreda	Šume	Voda
10	46	1205	198

Nadalje se skupina podataka za treniranje dijele proizvoljno na manje podskupine, tako da se skupine pojedinačno međusobno treniraju i preklapaju u algoritmu u pozadini. U ovom slučaju podaci se dijele na 10 skupina i koriste se za unakrsnu provjeru valjanosti (cross-validation CV). Nakon što su pronađeni optimalni odnosno najbolji parametri, konačni model je prilagođen cijelom skupu podataka za treniranje koristeći te parametre. Nadalje možemo provjeriti kako se ti konačni modeli ponašaju na skupu podataka za testiranje. To su metode koje će služiti za ispitivanje ispravnosti samog algoritma izrade predikcije (Valentin, 2024).

```
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. Metoda Random Forest (RF)

Koristeći *caret* paket kod trenira *Random Forest* model na podacima za treniranje (*dt\_train*) za predikciju klase (*class*) koristeći sve dostupne podatke. Za treniranje modela koristi se funkcija *caret::train*. Pretražuju se različite vrijednosti hiperparametra *mtry* kako bi se pronašla optimalna, a uključena je i procjena važnosti prediktorskih varijabli. Proces treniranja se kontrolira pomoću unakrsne provjere valjanosti definirane u objektu *ctrl*.

Funkcija *saveRDS* se koristi za spremanje modela u rds formatu kako bi se kasnije model učitao pomoću funkcije *readRDS* čime se izbjegava ponovno treniranje modela, a sprema se pod nazivom "model\_rfKANALI.rds". Linija *model\_rf* ispisuje sažetak treniranog modela.

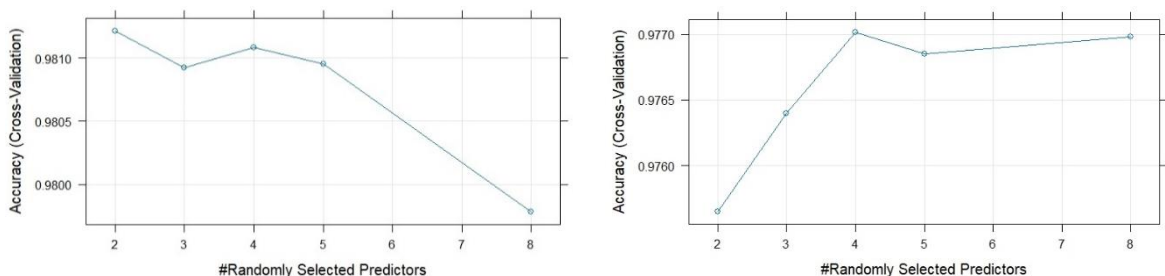
Funkcija *model\_rf\$times\$everything* prikazuje koliko je vremena bilo potrebno za različite dijelove procesa treniranja, uključujući vrijeme treniranja, predikcije i ukupno vrijeme.

Grafički prikaz dobiven naredbom *plot(model\_rf)* za satelitske snimke prije poplave prikazuje kako se najveća točnost postiže pri *mtry = 3 i 4*, čineći ih optimalnim vrijednostima za ovaj model, dok točnost opada pri višim vrijednostima *mtry*, što ukazuje na smanjenje performansi modela kada se uzima previše prediktora (slika 6.4.). Za satelitske snimke poslije poplave dobiven je grafički prikaz koji ukazuje na to da točnost raste pri svim *mtry* vrijednostima, osim malog pada pri *mtry = 4 i 5*.

```
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_rfKANALI.rds")
model_rf

model_rf$times$everything
plot(model_rf)
```



Slika 6.4. Grafički prikaz modela RF – prije i poslije poplave

Zatim se izračunava matrica konfuzije ( naredba *confusionMatrix*) za skup podataka za testiranje (*dt\_test*) koja uspoređuje definirane klase s predviđenim klasama modela. Tablica 6.5. prikazuje definirane i predviđene klase prije i poslije poplave. Model je najtočnije predvidio klase „Šume“ i „Voda“, dok su kod klase „Infrastruktura“ i „Poljoprivreda“ uočena određena odstupanja.

Uvidom u sveukupnu statistiku modela (tablica 6.6.) vidljivo je da je *Accuracy* i *Kappa* metrika veća za podatke prije poplave u odnosu na podatke poslije poplave. Točnost (*Accuracy*) izražava omjer broja ispravno klasificiranih uzoraka prema ukupnom broju uzoraka i trebala bi biti što je moguće veća. Razlog zašto se točnost obično kreće oko 99% je taj što se „unutarnja“ točnost često zadržava na visokom nivou jer se procjenjuje na podacima na kojima je algoritam treniran. Kada bi se koristili podaci na kojima algoritam nije bio treniran, dobila bi se „vanjska“ točnost koja bi bila znatno niža (npr. 75%). Na taj način se može preciznije ocijeniti koliko je predikcija zapravo uspješno napravljena (Medium-b, 2022; Medium-c, 2018). S druge strane Cohenova Kappa je mjera slaganja između dva zavisna kategorička uzorka i koristi se kada se želi znati jesu li mjerenja dvaju ocjenjivača u skladu (Statology, 2021).

Matrica pomaže u procjeni koliko je model dobro klasificirao podatke (tablica 6.7. i 6.8.). Linija *model\_rf\$finalModel* prikazuje detalje konačnog *Random Forest* modela nakon treniranja, a pomaže u razumijevanju koje značajke najviše doprinose klasifikaciji terena.

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

**Tablica 6.5.** Usporedba definiranih i predikcijskih klasa modela RF – prije i poslije poplave

Referenca / Predikcija	Infrastruktura		Poljoprivreda		Šume		Voda	
	Prije	Poslije	Prije	Poslije	Prije	Poslije	Prije	Poslije
Infrastruktura	5	9	0	0	0	3	0	0
Poljoprivreda	3	0	39	39	2	4	0	0
Šume	2	0	7	7	1203	1198	0	0
Voda	0	1	0	0	0	0	198	198

**Tablica 6.6.** Sveukupnu statistika modela RF – prije i poslije poplave

Ukupna statistika	Prije poplave	Poslije poplave
Accuracy	0.9904	0.9897
Kappa	0.9674	0.9655

**Tablica 6.7.** Statistika modela RF – prije poplave

Statistics by Class:	Class: Infrastruktura	Class: Poljoprivreda	Class: Šume	Class: Voda
Sensitivity	0.500000	0.84783	0.9983	1.0000
Specificity	0.100000	0.99646	0.9646	1.0000
Pos Pred Value	0.100000	0.88636	0.9926	1.0000
Neg Pred Value	0.996561	0.99505	0.9919	1.0000
Prevalence	0.006854	0.03153	0.8259	0.1357
Detection Rate	0.003427	0.02673	0.8245	0.1357
Detection Prevalence	0.003427	0.03016	0.8307	0.1357
Balanced Accuracy	0.750000	0.92214	0.9815	1.0000

**Tablica 6.8.** Statistika modela RF – poslije poplave

Statistics by Class:	Class: Infrastruktura	Class: Poljoprivreda	Class: Šume	Class: Voda
Sensitivity	0.900000	0.84783	0.9942	1.0000
Specificity	0.997930	0.99717	0.9724	0.9992
Pos Pred Value	0.750000	0.90698	0.9942	0.9950
Neg Pred Value	0.999309	0.99506	0.9724	1.0000
Prevalence	0.006854	0.03153	0.8259	0.1357
Detection Rate	0.006169	0.02673	0.8211	0.1357
Detection Prevalence	0.008225	0.02947	0.8259	0.1364
Balanced Accuracy	0.948965	0.92250	0.9833	0.9996

Nadalje funkcija *varImp* iz paketa *caret* izračunava važnost svake varijable u modelu *model\_rf*, a rezultati se pretvaraju u matricu. Matrica se zatim koristi za kreiranje toplinske mape (*heatmap* (1)), a boja prikazuje relativnu važnost svake varijable. Ukoliko boja ide više prema žutoj boji to ukazuje na lošiju klasifikaciju za određeni kanal (slika 6.5.). Funkcijom *importance* iz paketa *randomForest* prikazuje se važnost varijabli iz konačnog modela (*model\_rf\$finalModel*). Rezultati se ponovo pretvaraju u matricu i prikazuju grafikonom *heatmap* (2) (slika 6.6.). Zadnjom linijom koda dobivaju se grafikoni koji prikazuju koliko svaka varijabla doprinosi ukupnoj točnosti i kvaliteti konačnog modela (slika 6.7. i 6.8.).

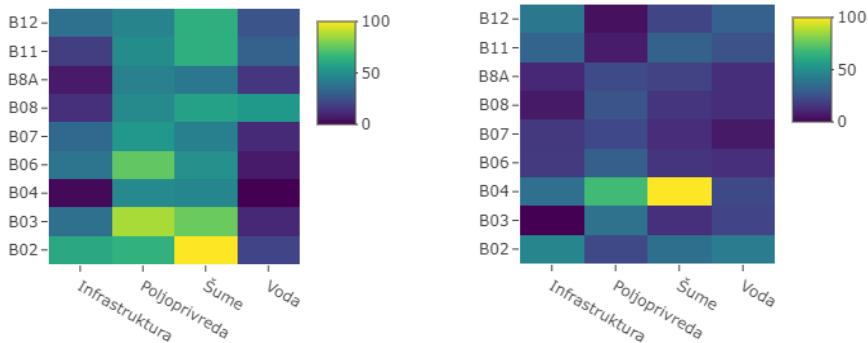
```

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

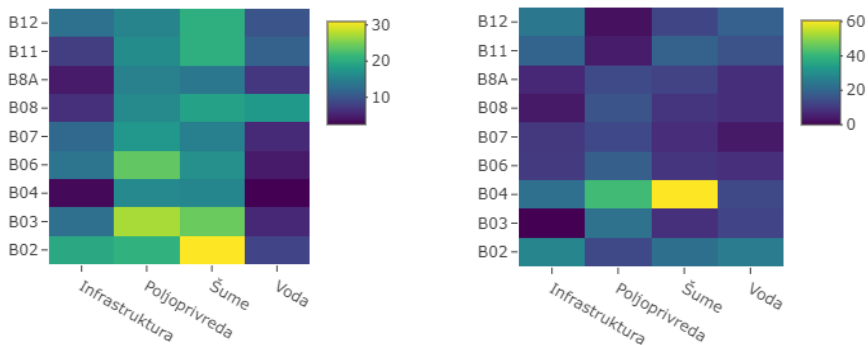
randomForest::importance(model_rf$finalModel) %>%
  .[, - which(colnames(.) %in% c("MeanDecreaseAccuracy",
                               "MeanDecreaseGini"))] %>%
  plot_ly(x = colnames(.), y = rownames(.), z = ., type = "heatmap",
          width = 350, height = 300)

randomForest::varImpPlot(model_rf$finalModel)

```

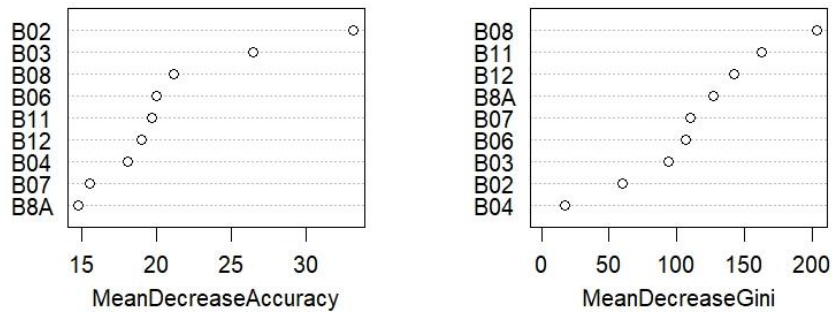


Slika 6.5. Prikaz grafikona heatmap (1) – prije i poslije poplave

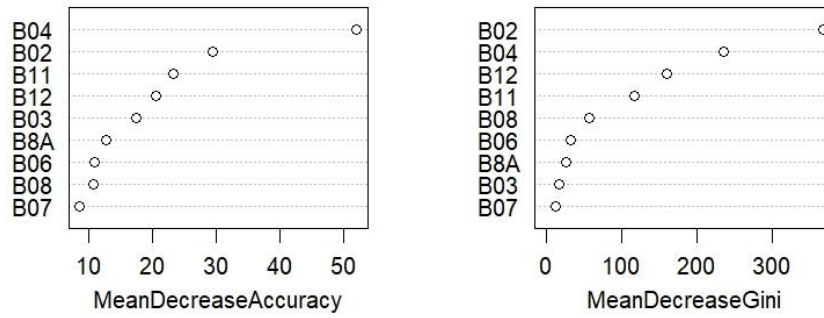


Slika 6.6. Prikaz grafikona heatmap (2) – prije i poslije poplave





Slika 6.7. Grafikon za brzu procjenu ključnih varijabli – prije poplave



Slika 6.8. Grafikon za brzu procjenu ključnih varijabli – poslije poplave

## 6.2. Metoda Support Vector Machine (SVM)

Sljedeća metoda strojnog učenja je *Support Vector Machine* gdje se model trenira na podacima za treniranje (*dt\_train*) za predikciju klase (*class*) koristeći linearnu funkciju *svmLinear3*. Kao i kod *RF* za treniranje modela koristi se funkcija *caret::train*. Naredba *expand.grid* stvara kombinacije hiperparametara za pretraživanje najbolje vrijednosti, a to su parametar *cost* sa tri vrijednosti (0.2, 0.5 i 1) i *Loss* sa dvije vrijednosti (L1 ili L2). Na taj način stvara se mreža (grid) sa šest kombinacija. Za poboljšanje parametara koristi se *svm\_grid* mreža, a proces treniranja se kontrolira pomoću unakrsne provjere valjanosti definirane u objektu *ctrl*.

Funkcijom *saveRDS* model se sprema u rds formatu pod nazivom "*model\_svmKANALI.rds*" i ispisuje se sažetak treniranog modela. Funkcija *model\_svm\$times\$everything* prikazuje koliko je vremena bilo potrebno za različite dijelove procesa treniranja, a naredbom *plot(model\_svm)* dobiva se grafički prikaz modela (slika 6.9.). Kao i kod modela *RF*, model *SVM* je najtočnije predvidio klase „Šume“ i „Voda“ (tablica 6.9.), dok je *Accuracy* i *Kappa* metrika veća za podatke prije poplave (tablica 6.10.). Na kraju se procjenjuje točnost modela na testnom skupu podatak za svaku klasu izračunavanjem matrice konfuzije (tablica 6.11. i 6.12.).

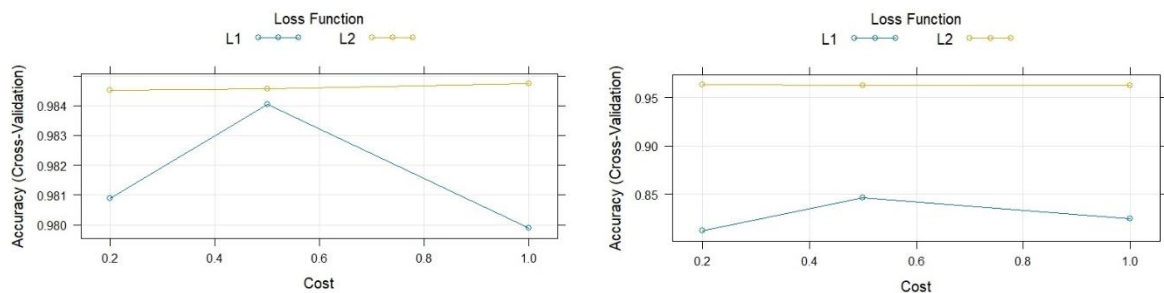
```
svm_grid <- expand.grid(cost = c(0.2, 0.5, 1),
                      Loss = c("L1", "L2"))

model_svm <- caret::train(class ~ ., method = "svmLinear3", data = dt_train,
                        allowParallel = TRUE,
                        tuneGrid = svm_grid,
                        trControl = ctrl)

saveRDS(model_svm, file = "model_svmKANALI.rds")
model_svm

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 6.9. Grafički prikaz modela SVM – prije i poslije poplave

Tablica 6.9. Usporedba definiranih i predikcijskih klasa modela SVM – prije i poslije poplave

Referenca / Predikcija	Infrastruktura		Poljoprivreda		Šume		Voda	
	Prije	Poslije	Prije	Poslije	Prije	Poslije	Prije	Poslije
Infrastruktura	7	9	1	0	0	0	0	0
Poljoprivreda	1	0	32	13	0	4	0	0
Šume	2	0	13	33	1205	1196	0	0
Voda	0	1	0	0	0	5	198	198

Tablica 6.10. Sveukupnu statistika modela SVM – prije i poslije poplave

Ukupna statistika	Prije poplave	Poslije poplave
Accuracy	0.9883	0.9705
Kappa	0.9599	0.8966

**Tablica 6.11.** Statistika modela SVM – prije poplave

Statistics by Class:	Class: Infrastruktura	Class: Poljoprivreda	Class: Šume	Class: Voda
Sensitivity	0.700000	0.69565	1.0000	1.0000
Specificity	0.999310	0.99929	0.9409	1.0000
Pos Pred Value	0.875000	0.96970	0.9877	1.0000
Neg Pred Value	0.997932	0.99018	1.0000	1.0000
Prevalence	0.006854	0.03153	0.8259	0.1357
Detection Rate	0.004798	0.02193	0.8259	0.1357
Detection Prevalence	0.005483	0.02262	0.8362	0.1357
Balanced Accuracy	0.849655	0.84747	0.9705	1.0000

**Tablica 6.12.** Statistika modela SVM – poslije poplave

Statistics by Class:	Class: Infrastruktura	Class: Poljoprivreda	Class: Šume	Class: Voda
Sensitivity	0.900000	0.28261	0.9925	1.0000
Specificity	1.000000	0.99717	0.8701	0.9952
Pos Pred Value	1.000000	0.76471	0.9731	0.9706
Neg Pred Value	0.999310	0.97712	0.9609	1.0000
Prevalence	0.006854	0.03153	0.8259	0.1357
Detection Rate	0.006169	0.00891	0.8197	0.1357
Detection Prevalence	0.006169	0.01165	0.8424	0.1398
Balanced Accuracy	0.950000	0.63989	0.9313	0.9976

### 6.3. Metoda Neural Network (NNET)

Kod posljednje metode definira se mreža (*nnet\_grid*) koja sadrži različite kombinacije parametar. Parametar *size* predstavlja broj neurona u skrivenom sloju (hidden layer), a parametar za regularizaciju *decay* služi za podešavanje modela. Model se trenira na podacima za treniranje koristeći sve dostupne podatke naredbom *train*. Naredbom *plot(model\_nnet)* dobiva se grafički prikaz modela (slika 6.10.). Model *SVM* je najtočnije predvidio klase „Šume“ i „Voda“ kao i prethodni modeli (tablica 6.13.), dok je *Accuracy* i *Kappa* metrika u ovom slučaju jednaka za podatke prije i poslije poplave (tablica 6.14.). Ponovnim računanjem matrice konfuzije procjenjuje se točnost modela na testnom skupu podatak (tablica 6.15. i 6.16.).

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

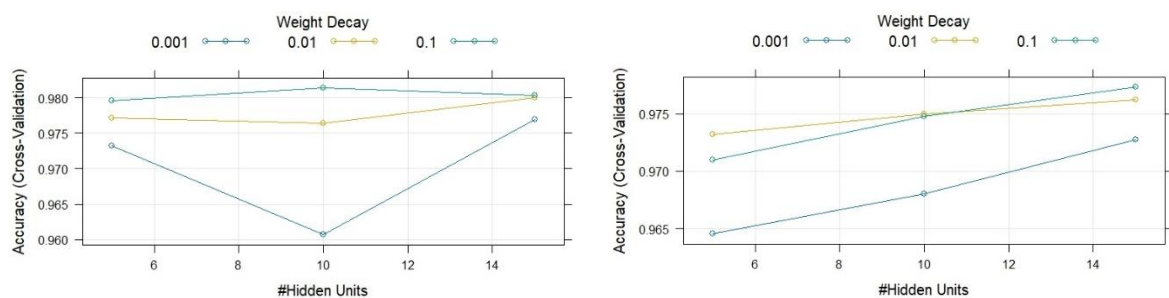
model_nnet <- train(class ~ ., method = 'nnet', data = dt_train,
                   importance = TRUE,
                   maxit = 1000,
                   allowParallel = TRUE,
                   tuneGrid = nnet_grid,
                   trControl = ctrl)

saveRDS(model_nnet, file = "model_nnetKANALI.rds")
model_nnet

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 6.10. Grafički prikaz modela NNET – prije i poslije poplave

**Tablica 6.13.** Usporedba definiranih i predikcijskih klasa modela NNET – prije i poslije poplave

Referenca / Predikcija	Infrastruktura		Poljoprivreda		Šume		Voda	
	Prije	Poslije	Prije	Poslije	Prije	Poslije	Prije	Poslije
Infrastruktura	9	8	3	0	0	0	0	0
Poljoprivreda	0	1	39	42	6	8	0	0
Šume	1	1	4	4	1199	1197	0	0
Voda	0	0	0	0	0	0	198	198

**Tablica 6.14.** Sveukupnu statistika modela NNET – prije i poslije poplave

Ukupna statistika	Prije poplave	Poslije poplave
Accuracy	0.9904	0.9904
Kappa	0.9679	0.968

**Tablica 6.15.** Statistika modela NNET – prije poplave

Statistics by Class:	Class: Infrastruktura	Class: Poljoprivreda	Class: Šume	Class: Voda
Sensitivity	0.900000	0.84783	0.9950	1.0000
Specificity	0.997930	0.99575	0.9803	1.0000
Pos Pred Value	0.750000	0.86667	0.9958	1.0000
Neg Pred Value	0.999309	0.99505	0.9765	1.0000
Prevalence	0.006854	0.03153	0.8259	0.1357
Detection Rate	0.006169	0.02673	0.8218	0.1357
Detection Prevalence	0.008225	0.03084	0.8252	0.1357
Balanced Accuracy	0.948965	0.92179	0.9877	1.0000

**Tablica 6.16.** Statistika modela NNET – poslije poplave

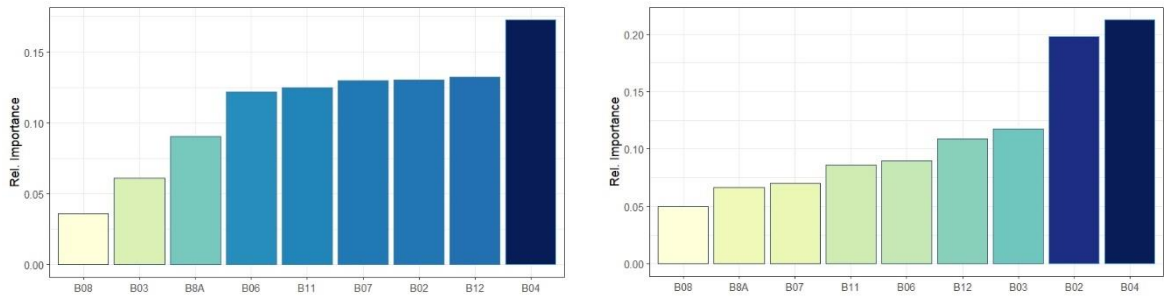
Statistics by Class:	Class: Infrastruktura	Class: Poljoprivreda	Class: Šume	Class: Voda
Sensitivity	0.800000	0.91304	0.9934	1.0000
Specificity	1.000000	0.99363	0.9803	1.0000
Pos Pred Value	1.000000	0.82353	0.9958	1.0000
Neg Pred Value	0.998622	0.99716	0.9689	1.0000
Prevalence	0.006854	0.03153	0.8259	0.1357
Detection Rate	0.005483	0.02879	0.8204	0.1357
Detection Prevalence	0.005483	0.03496	0.8239	0.1357
Balanced Accuracy	0.900000	0.95334	0.9868	1.0000

Na kraju se izvodi vizualizacija rezultata neuronske mreže i prikaz važnosti varijabli. Funkcija *colorRampPalette* iz paketa *grDevices* generira gradijentnu paletu boja, a pomoću funkcije *brewer.pal* se definira paleta koja sadrži 10 boja (*YlGnBu*). Prikaz važnosti varijabli u neuronskoj mreži izvodi se naredbom *garson(model\_nnet)* sa prilagođavanjem osi y i gradijenta boja grafa (slika 6.11.). Varijablom *cols\_rank\_import* dobiva se slika neuronske mreže gdje funkcija *rank* određuje poredak važnosti varijabli, a boje iz palete *cols* se raspoređuju prema važnosti varijabli (slika 6.12. i 6.13.). Graf i neuronska mreža su međusobno povezani bojama tako da tamnija boja predstavlja pouzdanije kanale u odnosu na kanale označene svjetlijom bojom.

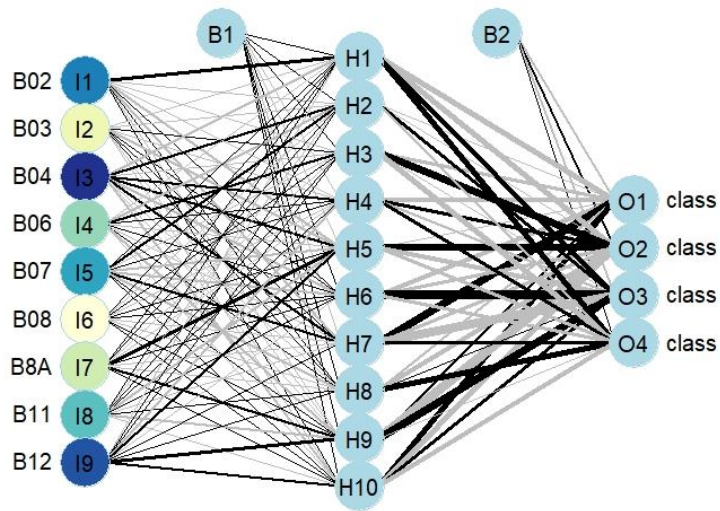
```
cols <- grDevices::colorRampPalette(colors = brewer.pal
                                   (n = 9, name = "YlGnBu"))(10)

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

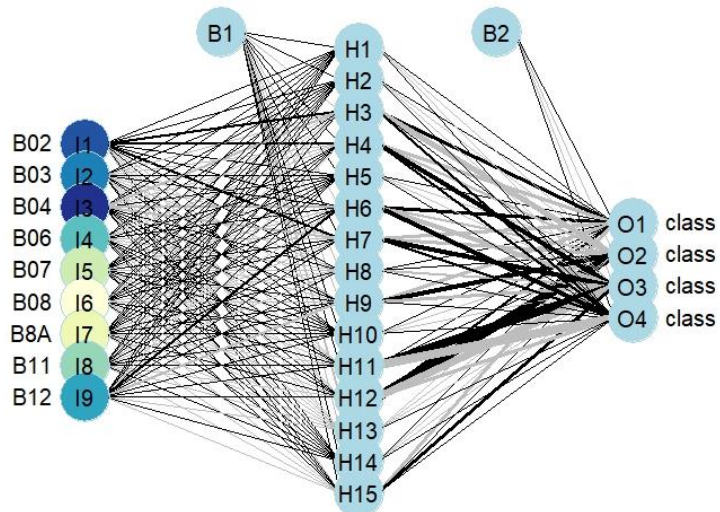
cols_rank_import <- cols[rank(garson(model_nnet, bar_plot = FALSE)$rel_imp)]
plotnet(model_nnet, circle_col = list(cols_rank_import, 'lightblue'))
```



Slika 6.11. Graf neuronske mreže – prije i poslije poplave



Slika 6.12. Neuronska mreža – prije poplave



Slika 6.13. Neuronska mreža – poslije poplave



#### 6.4. Usporedba modela RF, SVM i NNET

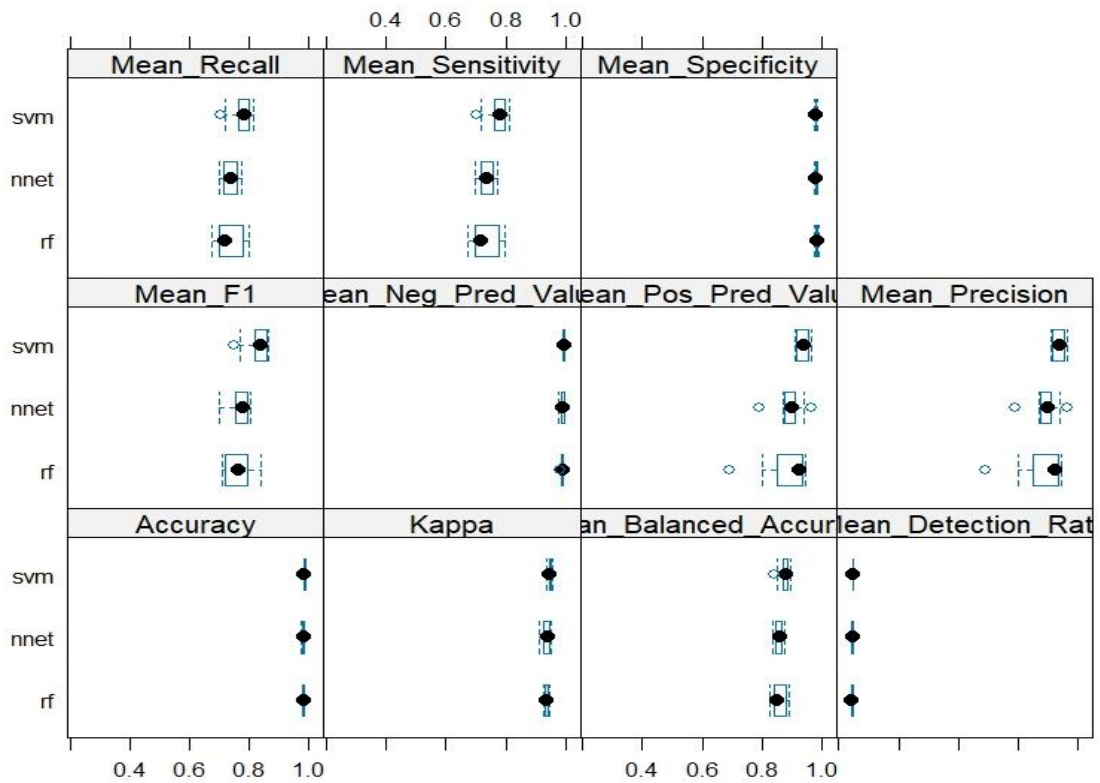
Modeli različitih algoritama (*Random Forest*, *Support Vector Machine* i *Neural Network*) spremaju se u listu pod varijablu `model_list` naredbom `list`. Nadalje se analiziraju njihove performanse i međusobno se statistički uspoređuju koristeći *resample* metode. Funkcija *resamples* iz paketa *caret* vrši usporedbu rezultata definiranih modela. Funkcija prikuplja metrike (*metrics*) kao na primjer *Accuracy*, *Kappa*, itd. za svaki model koristeći unakrsnu validaciju. Sa naredbom *bwplot* se kreiraju box plot-ovi koji prikazuje distribuciju svake metrike za svaki model na temelju rezultata funkcije *resamples*. Dobiveni grafikoni (slika 6.14. i slika 6.15.) daju uvid u promjenjivost performansi svakog modela i omogućava njihovu međusobnu usporedbu. Funkcija *diff* omogućuje usporedbu između modela za određenu metriku, u ovom slučaju za *Accuracy*, kako bi se provjerilo jesu li razlike statistički značajne, a pozivanjem varijable *t\_tests* se ispisuje tražena statistika. Na taj način mogu se statistički usporediti svi modeli na temelju svake metrike. Što su rezultati bliži vrijednosti 1, to su pouzdaniji. Na grafičkim prikazima se također može vidjeti raspršenost podataka. Ukoliko je rasipanje podataka usko (prikazano plavim kvadratom, linijom ili plavom točkom) oko srednje vrijednosti (crna točka) ukazuje na to da su rezultati i predikcije precizni i konzistentni. S druge strane, ako postoji veće rasipanje podataka, ali se i dalje može uočiti jasna srednja vrijednost, možemo zaključiti da model zadržava određenu razinu pouzdanosti u predikcijama. Na grafikonu za podatke prije poplave uočljivo je da *SVM* ima relativno najbolju preciznost srednje vrijednosti i najmanje raspršivanje podataka, dok *RF* i *NNET* također imaju relativno visoku pouzdanost i točnost bez obzira što imaju malo veću raspršivost. Za podatke poslije poplave dobiven je grafikon u kojem je drugačija situacija. Raspršenost podataka za svaki model je veća nego kog prvog grafikona, a *RF* ima najbolje performanse. Odabir modela s najboljim performansama ovisi o tome koje metrike smatramo najvažnijima, budući da svaki model ima svoje određene prednosti i mane.

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

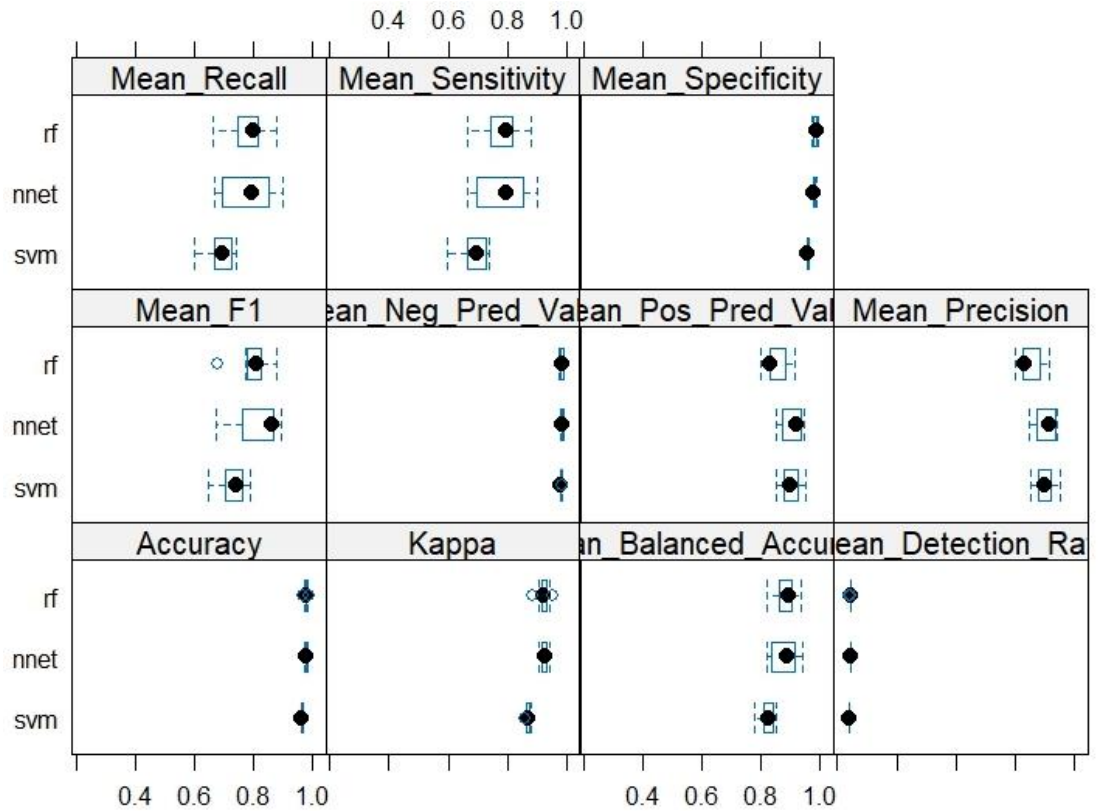
bwplot(resamples)

resamples$metrics

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



Slika 6.14. Grafikon za usporedbu modela RF, SVM i NNET – prije poplave



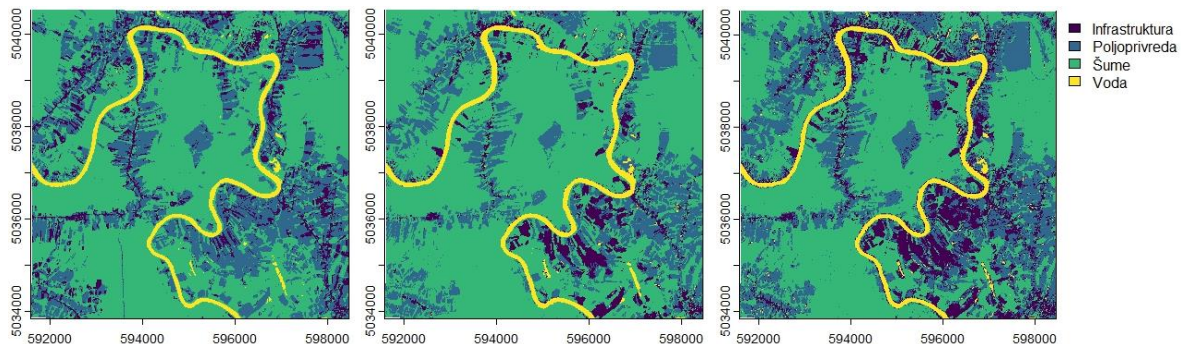
Slika 6.15. Grafikon za usporedbu modela RF, SVM i NNET – poslije poplave

Na posljetku se mjeri vrijeme potrebno za primjenu treniranih modela na rasterskim podacima odnosno na rasterskom sloju *s1*. Također se sve tri predikcije spremaju u tif formatu za daljnju obradu, analizu i vizualizaciju u programu ArcMap, a prikazane su na slici 6.12 i 6.13 dobivene naredbom *plot*.

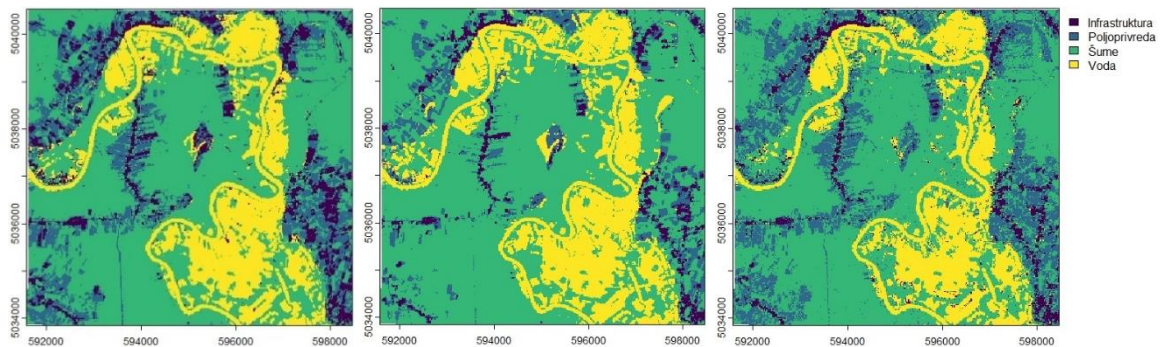
```
system.time({
  predict_rf <- raster::predict(object = s1,
                                model = model_rf, type = 'raw', na.rm=TRUE)
  predict_svm <- raster::predict(object = s1,
                                  model = model_svm, type = 'raw',na.rm=TRUE)
  predict_nnet <- raster::predict(object = s1,
                                   model = model_nnet, type = 'raw', na.rm=TRUE)}

plot(predict_rf)
plot(predict_svm)
plot(predict_nnet)

terra::writeRaster(predict_rf, "predict_rfKANALI.tif", overwrite = T)
terra::writeRaster(predict_svm, "predict_svmKANALI.tif", overwrite = T)
terra::writeRaster(predict_nnet, "predict_nnetKANALI.tif", overwrite = T)
```



Slika 6.16. Predikcije modela RF, SVM i NNET za satelitske snimke prije poplave



Slika 6.17. Predikcije modela RF, SVM i NNET za satelitske snimke poslije poplave

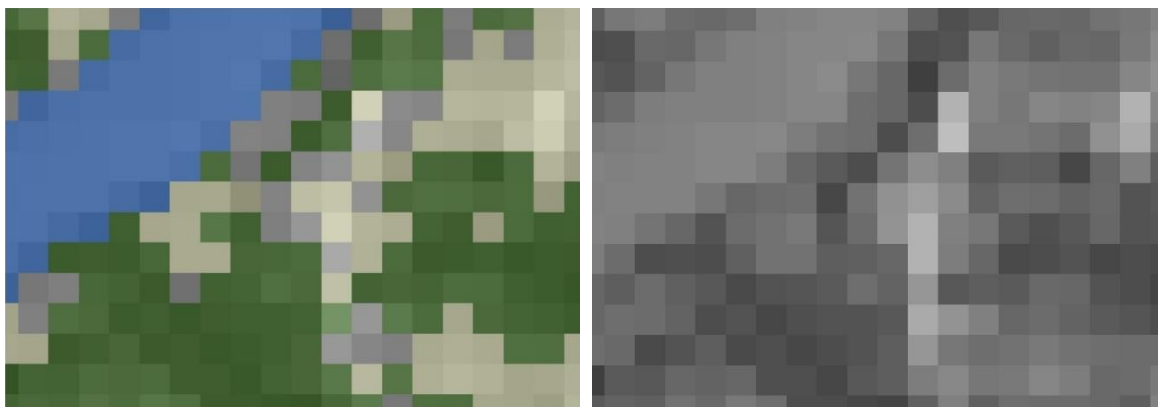
## 7. ANALIZA REZULTATA

Nakon spremanja predikcija koje su stvorene pomoću različitih modela, *Random Forest*, *Support Vector Machine* i *Neural Network*, u obliku rasterskih datoteka u tif formatu uvedene su u program ArcMap. U ArcMap-u su izdvojene klase, dobivene predikcijom, obojane prikladnim bojama radi lakšeg prepoznavanja. Klasi „Voda“ je pridružena tamno plava boja, klasi „Šume“ tamno zelena boja, klasi „Poljoprivreda“ svijetlo žuta boja, a klasi „Infrastruktura“ siva boja. Zatim su predikcije podešene na transparentnost od 80% kako bi se vizualno mogle usporediti sa Digitalnom ortofoto kartom (DOF) mjerila 1:1000 u pozadini. Digitalna ortofoto karta je preuzeta kao link sa servera Geoportal stranice te je uvedena u ArcMap povezivanjem na GIS server. S obzirom da je karta mjerila 1:1000 omogućuje prikaz objekata veličine 1 metra, a prostorna rezolucija joj je otprilike 0,1 do 0,5 metara po pikselu.

Preklapanjem predikcija sa DOF kartom i satelitskim snimkama vidljiva je razlika u rezoluciji na slikama 7.1. i 7.2. Zbog manje rezolucije satelitskih snimaka u odnosu na veću rezoluciju DOF karte može se vidjeti koliko rezolucija utječe na točnost klasifikacije terena. Također se može uočiti da zbog rezolucije satelitskih snimaka od 20 m, pikseli zauzimaju veću površinu svake klase. Pikseli zauzimaju veću površinu pogotovo za klasu „Infrastruktura“ nego što je u stvarnosti jer su površine infrastrukture uglavnom manje od 20x20m.



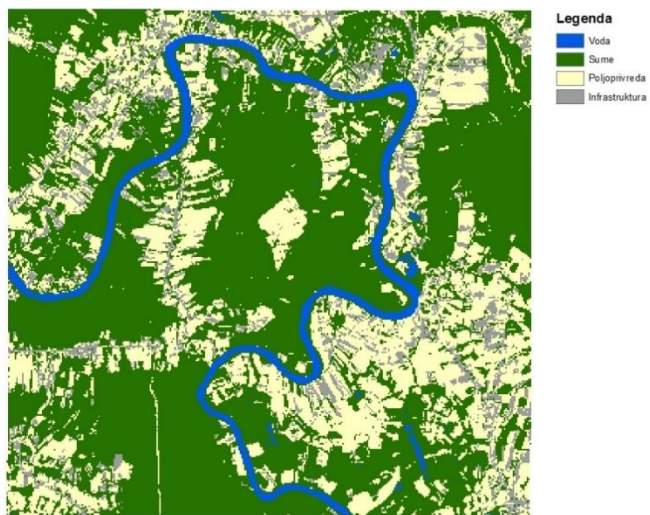
**Slika 7.1.** Digitalna ortofoto karta (DOF) preklapljen sa predikcijom



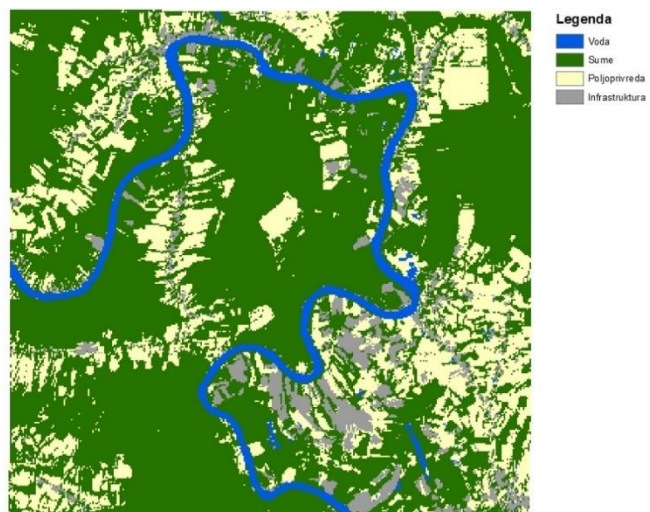
**Slika 7.2.** Satelitska snimka preklopljena sa i bez predikcije

Međutim, bez obzira na rezoluciju može se reći da predikcije dobivene modelima *Random Forest*, *Support Vector Machine* i *Neural Network* vizualno zadovoljavaju očekivanja. Preklapanjem se generalno uočava bolja klasifikacija klasa „Šume“ i „Voda“ kod svih triju predikcija, dok kod klasa „Infrastruktura“ i „Poljoprivreda“ ima određenih odstupanja. Takva odstupanja su utvrđena i statistikom u poglavlju 6. „Obrada podataka“ gdje predikcije pridodaju klasu osim definiranoj klasi i drugim klasama. Uvidom u DOF kartu može se primijetiti kako su boje piksela za klase „Infrastruktura“ i „Poljoprivreda“ vrlo slične pa vjerojatno dolazi do „zbunjivanja“ algoritma pa određeni dio predikcijskih klasa pridodaje krivoj klasi.

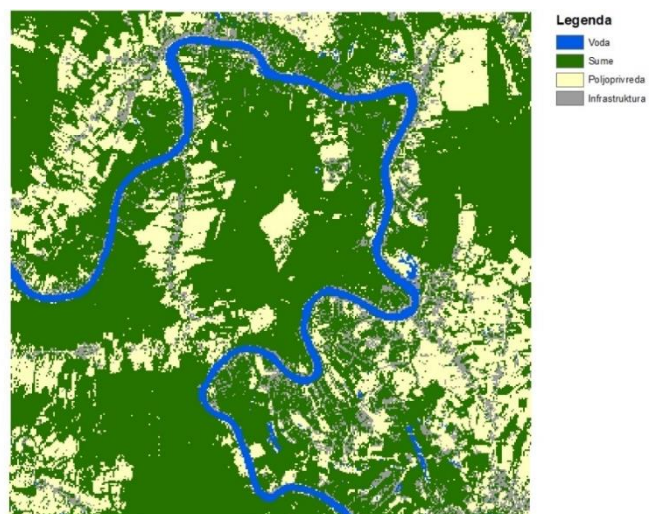
Vizualnim uspoređivanjem predikcija prije poplave sa DOF kartom u pozadini može se zaključiti da model *RF* i *NNET* daju najbolje rezultate, a svaki ima svoje prednosti i mane. Sve klase su otprilike dobro raspoređene bez obzira na određena odstupanja koja su prethodno napomenuta. Generalno je kod modela *RF* (slika 7.3.) najbolja raspodjela klase „Poljoprivreda“, a kod modela *NNET* (slika 7.5.) raspodjela klase „Infrastruktura“. Model *RF* je najbolje predvidio klasu „Poljoprivreda“, međutim na određenim mjestima ju je klasificirao kao klasu „Infrastruktura“ što ne odgovara stvarnosti. Kod modela *NNET* najbolja je raspodjela klase „Infrastruktura“ koja odgovara stvarnosti međutim nije na mnogo mjesta prepoznao klasu „Poljoprivreda“ te ju je pridodao klasi „Šume“. Nadalje, lošije rezultate daje predikcija dobivena modelom *SVM*, a to se najbolje primjećuje u donjem desnom kutu slike 7.4. gdje model veliku površinu terena pridodaje klasi „Infrastruktura“ umjesto klasi „Poljoprivreda“ što odgovara stvarnosti.



Slika 7.3. Predikcija pomoću modela RF – prije poplave

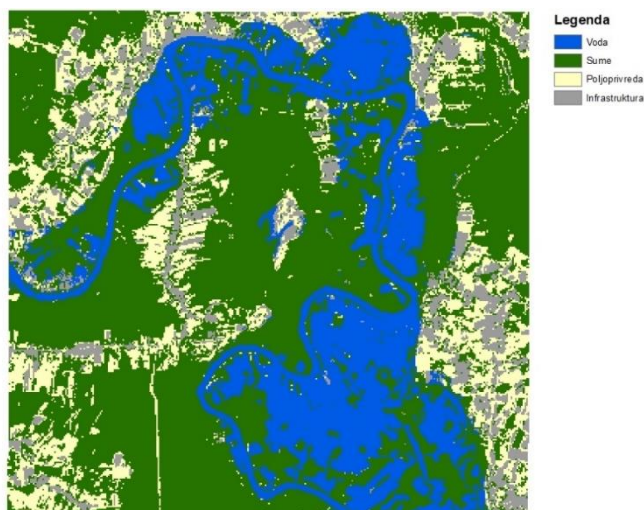


Slika 7.4. Predikcija pomoću modela SVM – prije poplave

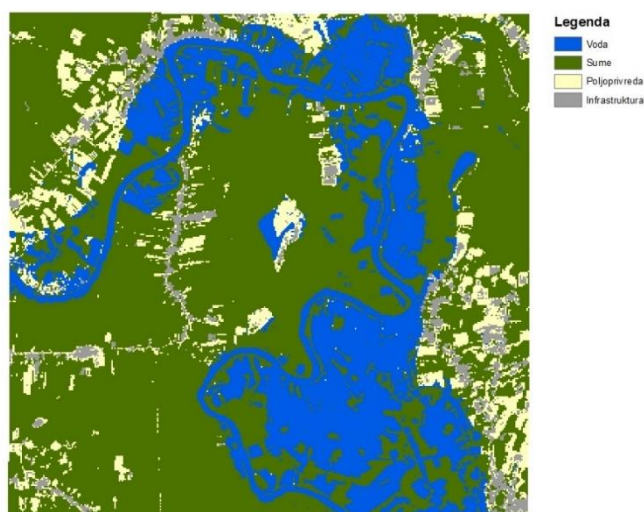


Slika 7.5. Predikcija pomoću modela NNET – prije poplave

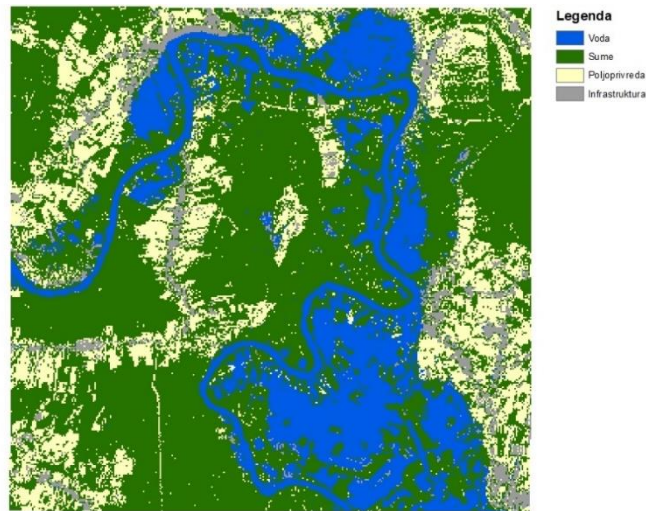
Što se tiče predikcija na satelitskim snimkama poslije poplave može se zaključiti kako su predikcijskom klasifikacijom vodenih površina dobiveni izvrsni rezultati, a prikazane su na slikama 7.6., 7.7. i 7.8. Međutim, klasifikacija terena za ostale klase je dosta različita s obzirom na klasifikaciju terena prije poplave. Puno je lošija jer klasa „Infrastruktura“ i „Poljoprivreda“ za sve tri predikcije ne odgovaraju stvarnosti, a najlošiji rezultati su dobiveni modelom *SVM* gdje klase „Infrastruktura“ i „Poljoprivreda“ prepoznaje i klasificira kao klasu „Šume“.



**Slika 7.6.** Predikcija pomoću modela RF – poslije poplave



**Slika 7.7.** Predikcija pomoću modela SVM – poslije poplave

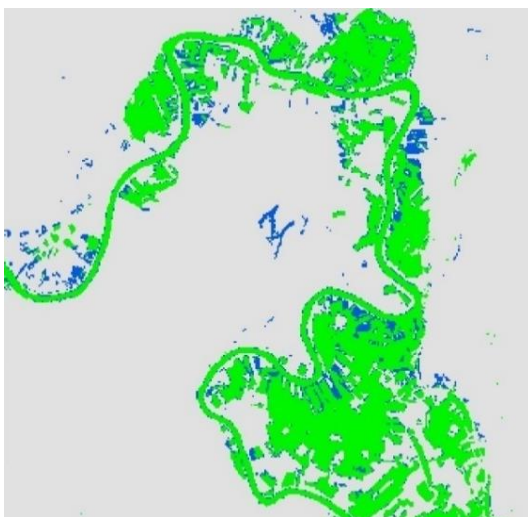


**Slika 7.8.** Predikcija pomoću modela NNET – poslije poplave

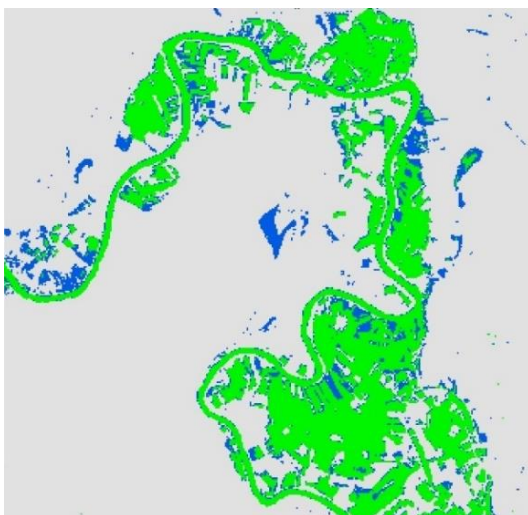
Nadalje, predikcije za sva tri modela su uspoređene sa NDWI indeksom normalizirane razlike vode. NDWI (Normalized Difference Water Index) je indeks koji se koristi u analizi satelitskih snimaka za identifikaciju i praćenje površinskih voda, poput rijeka, jezera i drugih vodenih tijela. Pomaže u razlikovanju voda od ostalih objekata na Zemljinoj površini, poput vegetacije i tla (EOS Data Analytics, 2021). NDWI indeks se dobiva tako da se najprije iz satelitskih snimaka izdvoje vrijednosti refleksije svjetlosti u zelenom i blisko-infracrvenom (NIR) spektru (kanali 3 i 5). Zatim se oduzme refleksija u NIR pojasu od refleksije u zelenom pojasu, a rezultat se podijeli s njihovim zbrojem. Indeks je dobiven računanjem u R programskom jeziku, a spremljen je kao tif datoteka kako bi se mogao preklapati sa predikcijama u ArcMap programu. Slike 7.9., 7.10. i 7.11. prikazuju preklapanje predikcija modela *RF*, *SVM* i *NNET* (tamno plava boja) sa NDWI indeksom (zelena boja).

Usporedbom NDWI indeksa s predikcijama svih triju modela može se zaključiti da se indeks najbolje preklapa sa modelom *NNET* (slika 7.11.). Modeli *RF* (slika 7.9.) i *SVM* (slika 7.10.) se velikim dijelom preklapaju sa NDWI indeksom, ali prikazuju poplavljene površine i na novim područjima. Budući da je NDWI indeks dobiven korištenjem dva kanala, a predikcije korištenjem rasterskog sloja koji se sastoji od devet kanala, ne može se sa sigurnošću reći koja metoda daje najbolje rezultate. Budući da je za strojno učenje korišteno više kanala, može se pretpostaviti da daje bolje rezultate u odnosu na NDWI indeks. To je samo pretpostavka te bi se dobiveni rezultati u oba slučaja trebali uvijek usporediti, provjeriti i potvrditi s terenskim istraživanjima. Međutim, bez obzira na to napravljene su usporedbe tih dviju metoda sa satelitskom snimkom poslije poplave (slika 7.12.).

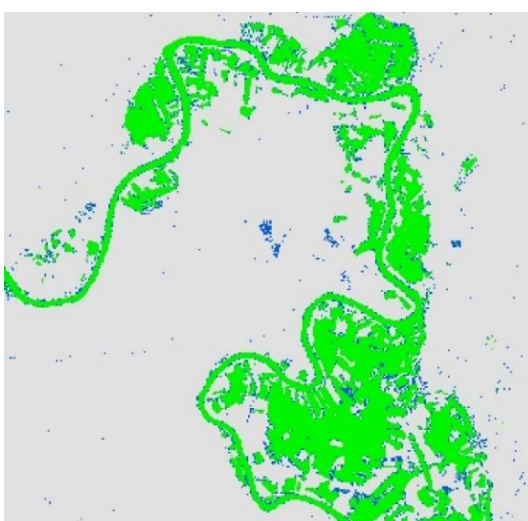




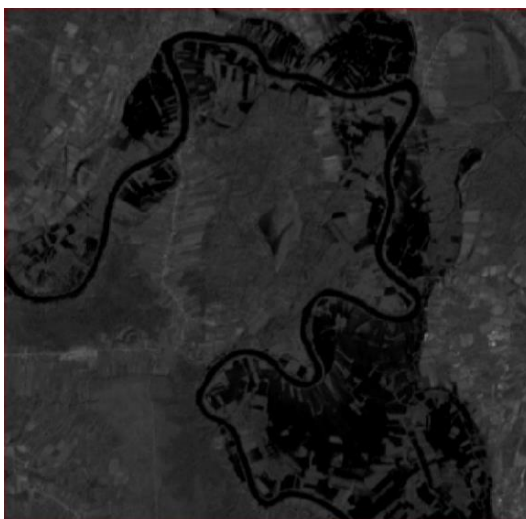
**Slika 7.9.** Predikcija modela RF preklopljena sa NDWI indeksom



**Slika 7.10.** Predikcija modela SVM preklopljena sa NDWI indeksom



**Slika 7.11.** Predikcija modela NNET preklopljena sa NDWI indeksom



**Slika 7.12.** Satelitska snimka poslije poplave

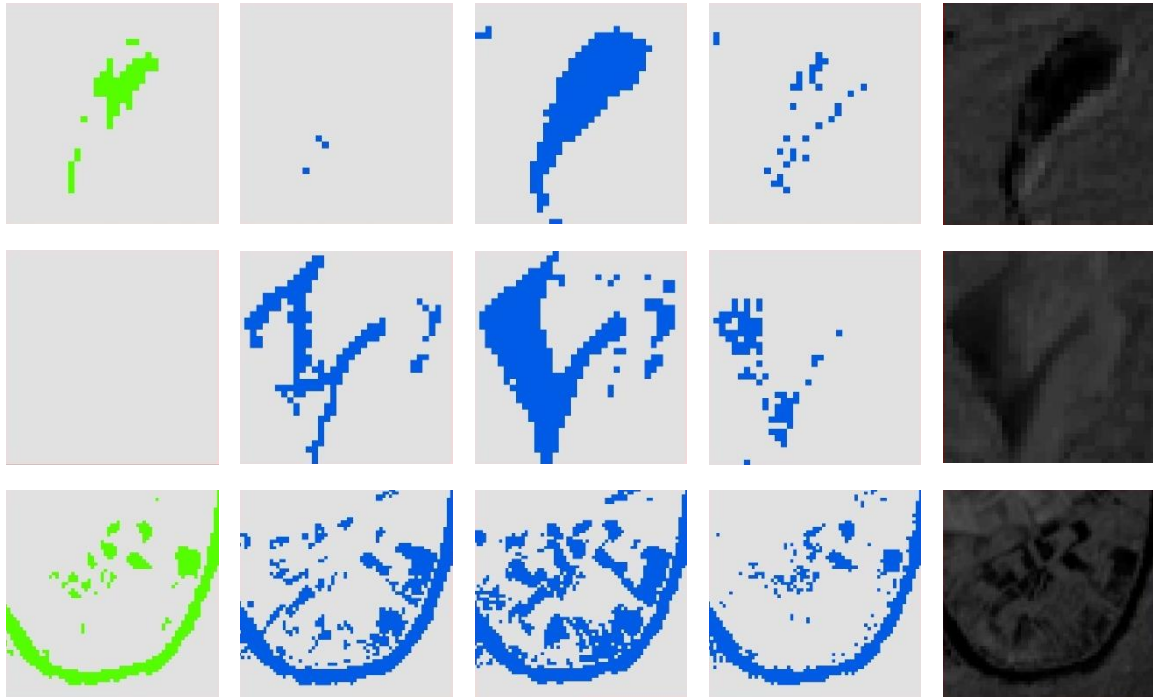
Uvidom u par uvećanih uzoraka može se utvrditi sljedeće. U prvom redu slike 7.13. vidljivo je da se predikcija dobivena modelom *SVM* najbolje preklapa sa NDWI indeksom, ali prepoznaje nešto veću površinu kao klasu "Voda" u odnosu na NDWI indeks. S obzirom na satelitsku snimku može se pretpostaviti da metoda *SVM* daje najbolje rezultate. Predikcije dobivene modelima *RF* i *NNET* gotovo uopće ne prepoznaju vodenu površinu.

U drugom redu slike 7.13. može se vidjeti da NDWI indeks nije prepoznao vodenu površinu, dok sve predikcije jesu. Uvidom u satelitsku snimku postoji sumnja je li to vodena površina s obzirom da nije jako izražena kao u prethodnom uzorku. Eventualno se može pretpostaviti da se radi o zaostaloj maloj količini vode, pa iz tog razloga nije jako izražena na satelitskoj snimci koja je preuzeta nekoliko dana poslije velike poplave kada se većina vode povukla. Zbog te činjenice postoji mogućnosti da NDWI daje neprecizne rezultate te da ne prepoznaje male količine vode. Predikcija dobivena modelom *SVM* klasificira nešto veću površinu kao klasu "Voda" u odnosu na predikcije modela *RF* i *NNET*.

U zadnjem redu slike 7.13. se ponovo uočava da predikcije klasificiraju veću vodenu površinu u odnosu na NDWI indeks. Pretpostavlja se da predikcije klasificiraju veću vodenu površinu iz istog razloga kao i u prethodnom uzorku gdje su na satelitskoj snimci vidljive preostale poplavljene površine manje izraženosti što indeks ne prepoznaje. Jedino se predikcija modela *NNET* gotovo potpuno preklapa sa NDWI indeksom i ne klasificira veću vodenu površinu kao predikcije modela *RF* i *SVM*. Predikcija dobivena modelom *RF* se gotovo uopće ne preklapa sa indeksom, a klasificira ostale površine kao klasu „Voda“.

Predikcija modela *SVM* se preklapa sa NDWI indeksom, ali klasificira puno veću vodenu površinu u odnosu na predikciju modela *RF*.

Generalno gledajući cijelo područje istraživanja, predikcije modela *RF* i *SVM* klasificiraju nešto veću vodenu površinu u odnosu na NDWI indeks, dok se predikcija modela *NNET* gotovo potpuno preklapa sa indeksom.



**Slika 7.13.** Usporedba NDWI indeksa, predikcija modela RF, SVM, NET i satelitske snimke poslije poplave

## 8. ZAKLJUČAK

Analiza rezultata odnosno predikcija prije i poslije poplave pokazuje da modeli *Random Forest (RF)* i *Neural Network (NNET)* pružaju zadovoljavajuće rezultate, ali sa određenim prednostima i manama. Prije poplave, *RF* model najbolje klasificira klasu „Poljoprivreda“, dok *NNET* model preciznije klasificira klasu „Infrastruktura“. Model *Support Vector Machine (SVM)* pokazuje slabije performanse, posebno u razlikovanju klasa „Poljoprivreda“ i „Infrastruktura“. Nakon poplave, usporedba predikcija s NDWI indeksom otkriva da model *NNET* postiže najtočnije rezultate u prepoznavanju površinskih voda, dok *RF* i *SVM* modeli klasificiraju nešto veću površinu kao klasu „Voda“, ali su te razlike unutar prihvatljivih granica. Također, u svim predikcijama nakon poplave se primjećuje lošija klasifikacija klasa „Infrastruktura“ i „Poljoprivreda“ u odnosu na situaciju prije poplave, a značajno lošije rezultate daje model *SVM* jer klase „Infrastruktura“ i „Poljoprivreda“ često pogrešno klasificira kao „Šume“. Ovaj problem upućuje na izazove u prepoznavanju tih klasa u uvjetima promijenjenim poplavom.

Izazovi u klasifikaciji proizlaze isto tako iz rezolucije satelitskih snimaka i sličnosti boja piksela, što dovodi do određenih odstupanja u predikcijama. Unatoč tim ograničenjima, predikcije se generalno slažu s očekivanjima i pružaju korisne informacije za daljnje analize.

Analiza pokazuje da i uz ograničenu količinu ulaznih podataka, u ovom slučaju 27 nasumično odabranih poligona, strojno učenje može pružiti zadovoljavajuće rezultate u klasifikaciji terena. Iako se radi o relativno malom uzorku podataka, rezultati su korisni i pružaju solidnu osnovu za daljnje analize. Ova činjenica naglašava fleksibilnost i prilagodljivost modela strojnog učenja, budući da se ulazni podaci mogu kontinuirano nadograđivati kako bi se poboljšala preciznost klasifikacije, posebno za klase „Infrastruktura“ i „Poljoprivreda“. U tom kontekstu, buduće nadogradnje kroz povećanje količine i kvalitete podataka vjerojatno bi dovele do značajnog poboljšanja u predikcijama ovih klasa. Takav pristup obradi podataka rezultirao bi preciznijim klasifikacijama i smanjenjem grešaka koje su primijećene u sadašnjim rezultatima.

Osim tehničkih aspekata, predikcije klasifikacije terena za vrijeme poplave imaju praktičnu primjenu. Na temelju tih podataka moguće je izraditi karte koje bi značajno unaprijedile učinkovitost civilne zaštite. Takve karte mogu identificirati najugroženija područja i pomoći u planiranju odgovarajućih mjera zaštite, evakuacije i sanacije.

Posebno korisno može biti određivanje površine poplavljenih područja pomoću klase „Voda“, što se može izračunati na temelju broja piksela pridruženih toj klasi.

Kontinuirano unapređenje modela sa novim podacima može dovesti do preciznijih predikcija i, posljedično, do praktičnih rješenja za zaštitu i upravljanje u kriznim situacijama poput poplava.

## 9. LITERATURA

MORAGA, P. 2023. *Spatial Statistics for Data Science: Theory and Practice with R*. Chapman & Hall/CRC Data Science Series.

URL: <https://www.paulamoraga.com/book-spatial/index.html> (25.7.2024.)

WICKHAM, H., GROLEMUND, G. 2017. *R for Data Science*

URL: <https://r4ds.hadley.nz/> (25.7.2024.)

ŠARIĆ, G. 2022. *Određivanje poplavljenog područja korištenjem multispektralnih satelitskih snimaka u R programskom sustavu*. Diplomski rad. Zagreb: Rudarsko-geološko-naftni fakultet.

SZILLER, L. 2023. *Klasifikacija litologije dijela Sjevernog Velebita korištenjem multispektralnih satelitskih snimki Sentinel-2 u programskom jeziku R*. Diplomski rad. Zagreb: Rudarsko-geološko-naftni fakultet.

Telegram, 2023. *Poplave u Hrvatskoj*.

URL: <https://www.telegram.hr/> (15.7.2024.)

DHMZ-a, 2024. *Vodostaji rijeka i dalje visoki*.

URL: [https://meteo.hr/objave\\_najave\\_natjecaji.php?section=onn&param=objave&el=priopcenja&daj=pr17052023hidro](https://meteo.hr/objave_najave_natjecaji.php?section=onn&param=objave&el=priopcenja&daj=pr17052023hidro) (15.7.2024.)

DHMZ-b, 2024. *Hidrološke postaje*.

URL: <https://hidro.dhz.hr/> (15.7.2024.)

Copernicus-1, 2024. *About Copernicus*.

URL: <https://www.copernicus.eu/en/about-copernicus> (22.7.2024.)

Copernicus-2, 2024. *Copernicus Programme*.

URL: <https://sentiwiki.copernicus.eu/web/copernicus-programme> (22.7.2024.)

Sentinel-a, 2024. *S2 Mission*

URL: <https://sentiwiki.copernicus.eu/web/s2-mission> (24.7.2024.)

Sentinel-b, 2024. *Introducing Sentinel-2*.

URL: [https://www.esa.int/Applications/Observing\\_the\\_Earth/Copernicus/Sentinel-2/Introducing\\_Sentinel-2](https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2/Introducing_Sentinel-2) (24.7.2024.)

Copernicus Browser-1, 2024. *About the Browser*.

URL: <https://documentation.dataspace.copernicus.eu/Applications/Browser.html> (24.7.2024.)

Copernicus Browser-2, 2024. *Copernicus Browser*.

URL: <https://browser.dataspace.copernicus.eu/> (24.7.2024.)

Kent State University, 2024. *Statistical & qualitative data analysis software: about R and RStudio*.

URL: <https://libguides.library.kent.edu/statconsulting/free-software> (25.7.2024.)

Patwal, M., 2020. *A Short Introduction to R Programming*.

URL: <https://bookdown.org/manishpatwal/bookdown-demo/r-studio-overview.html#console-pane> (25.7.2024.)

R package Wikipedia, 2024. *R package*.

URL: [https://en.wikipedia.org/wiki/R\\_package](https://en.wikipedia.org/wiki/R_package) (25.7.2024.)

posit™ Support, 2024. *Quick list of useful R packages*.

URL: <https://support.rstudio.com/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages> (25.7.2024.)

RDocumentation, 2024. *Search all 28.768 R packages on CRAN and Bioconductor*.

URL: <https://www.rdocumentation.org/> (25.7.2024.)

ESRI-1, 2024. *ArcGIS: Esri's enterprise geospatial platform*.

URL: <https://www.esri.com/en-us/arcgis/geospatial-platform/overview> (30.7.2024.)

ESRI-2, 2024. *WHAT IS GIS? Geographic Information Systems (GIS)*.

URL: <https://www.esri.com/en-us/what-is-gis/overview#choosing-the-right-location>

(30.7.2024.)

ArcGIS Desktop, 2024. *What is ArcMap?*.

URL: <https://desktop.arcgis.com/en/arcmap/latest/map/main/what-is-arcmap-.htm>

(30.7.2024.)

Medium-a, 2023. *Getting Started with Geographic Information Systems (GIS): A Comprehensive Guide*.

URL: <https://medium.com/> (30.7.2024.)

GIS, 2024. *Remote Sensing . GIS . Maps: Knowledge Resource for Remote Sensing, GIS, GPS and Advancements in Geomatics*.

URL: <https://rsgislearn.blogspot.com/2007/05/vector-data-vs-raster-data.html> (30.7.2024.)

City of Newberg, 2024. *Geographic Information System (GIS)*.

URL: <https://www.newbergoregon.gov/engineering/page/about-newberg-gis> (30.7.2024.)

Javapoint, 2021. *Random Forest Algorithm*.

URL: <https://www.javatpoint.com/machine-learning-random-forest-algorithm> (2.8.2024.)

UC Business Analytics R Programming Guide, 2024. *Support Vector Machine*.

URL: <https://uc-r.github.io/svm> (2.8.2024.)

Datacamp, 2018. *Support Vector Machines in R*.

URL: <https://www.datacamp.com/> (2.8.2024.)

Geeksforgeeks, 2023. *Neural Networks Using the R nnet Package*.

URL: <https://www.geeksforgeeks.org/neural-networks-using-the-r-nnet-package/>

(2.8.2024.)

Packtpub, 2024. *Pros and cons of neural networks*.

URL: <https://subscription.packtpub.com/> (2.8.2024.)



Valentin, S. 2024. *R - using Random Forests, Support Vector Machines and Neural Networks for a pixel based supervised classification of Sentinel-2 multispectral images.*

URL: <https://valentinitnelav.github.io/> (5.6.2024.)

Medium-b, 2022. *Data Science: Machine Learning Models Metrics.*

URL: <https://medium.com/@i.cemozcelik/data-science-machine-learning-models-metrics-77f9f77c2ff4> (5.6.2024.)

Medium-c, 2018. *Understanding Confusion Matrix.*

URL: <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62> (5.6.2024.)

Statology, 2021. *Cohen's Kappa Statistic: Definition & Example.*

URL: <https://www.statology.org/cohens-kappa-statistic/> (5.6.2024.)

EOS Data Analytics, 2021. *Normalized Difference Water Index (NDWI).*

URL: <https://eos.com/make-an-analysis/ndwi/> (5.6.2024.)