

Klasifikacija zemljišnog pokrova iz podataka Sentinel satelitskih misija

Dino Dobrinić, mag. ing. geod. et geoinf.¹, prof. dr. sc. Damir Medak², Ana Katanec, univ. bacc. ing. geod. et geoinf.³

¹Geodetski fakultet, Sveučilište u Zagrebu, Kačićeva 26, Zagreb, Hrvatska, ddobrinic@geof.hr

²Geodetski fakultet, Sveučilište u Zagrebu, Kačićeva 26, Zagreb, Hrvatska, dmedak@geof.hr

³Geodetski fakultet, Sveučilište u Zagrebu, Kačićeva 26, Zagreb, Hrvatska, akatanec@geof.hr

SAŽETAK: Cilj ovog istraživanja je klasifikacija zemljišnog pokrova iz snimaka prikupljenih pomoću Sentinel-1 (S1) i Sentinel-2 (S2) satelita. Copernicus je EU program usmjeren ka razvoju Europskog informacijskog servisa zasnovanog na satelitima za opažanje Zemlje, podacima s kopna i drugih doprinosećih misija. U sklopu programa, između ostalog, bez naknade dostupne su svim korisnicima radarske (S1) i optičke (S2) satelitske snimke visoke rezolucije. Radar posjeduje vlastiti izvor zračenja pa se snimanje može odvijati neovisno o dobu dana te zahvaljujući duljim valnim duljinama, neovisan je o vremenskim prilikama. Sentinel-1 sateliti lansirani su 2014. i 2016. godine, i na njima se nalazi aktivni radar (engl. Synthetic Aperture RADAR - SAR) te snimke su dostupne svakih 6 dana. Sentinel-2 je multispektralna satelitska misija (sastoji se od satelita S-2A i S-2B) visoke rezolucije koja opaža Zemljinu površinu u 13 spektralnih kanala prostornih rezolucija od 10 m, 20 m i 60 m.

U ovom radu usporediti će se radarske i optičke snimke jednake prostorne rezolucije od 10 m pomoću različitih metoda strojnog učenja. Uz Random Forest (RF) metodu, koji se najčešće koristi za nadziranu klasifikaciju, koristiti će se još XGBoost i Support Vector Machine (SVM) metode. Pritom, RF i XGBoost metode kombiniraju predviđanja više modela (engl. ensemble learning), dok SVM definira hiperravninu koja klasificira sve vektore iz trening skupa podataka u dvije ili više klase. Glavni cilj ovog istraživanja je usporediti rezultate klasifikacije zemljišnog pokrova prema pojedinim klasama za radarske i optičke snimke te odrediti mogućnosti njihovih primjena u daljnjim istraživanjima.

Ključne riječi: klasifikacija, Random Forest, Sentinel, Support Vector Machine, XGBoost.

1. UVOD

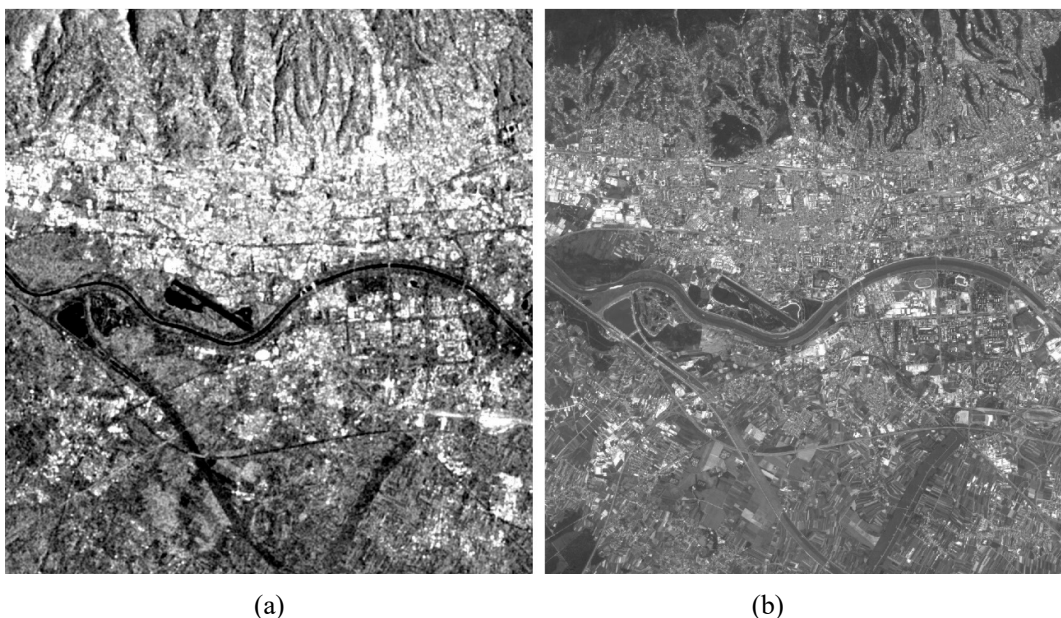
Daljinska istraživanja se zasnivaju na korištenju spektra elektromagnetskih valova, koji prenose informacije od objekta do na njih osjetljivog uređaja (senzora), koji ih prima i bilježi. Zbog mogućnosti prikupljanja velike količine prostornih podataka u vrlo kratkom vremenu, metodologija daljinskih istraživanja sve se intenzivnije povezuje sa znanjima i vještinama iz drugih znanstvenih disciplina (Valozić, 2014). Upotreba satelitskih snimaka za klasifikaciju zemljišnog pokrova počela se koristiti početkom 1970-ih godina, kada su Sjedinjene Američke Države lansirale satelit Landsat 1 (Mulla, 2013). Napretkom tehnologije, a i povećanjem potreba za korištenjem daljinskih istraživanja u globalnim aplikacijama, omogućen je razvoj novih i sofisticiranijih sustava za praćenje Zemljine površine. Glavni trendovi prilikom razvoja sustava su: rezolucija (prostorna, spektralna, vremenska i radiometrijska), brzina obrade podataka i mogućnosti analize prostornih podataka.

Copernicus je program Europske Unije usmjeren ka razvoju Europskog informacijskog servisa zasnovanog na satelitima za opažanje Zemlje, podacima s kopna i drugih doprinosećih misija. U sklopu programa, između ostalog, bez naknade dostupne su svim korisnicima radarske (Sentinel-1; S1) i optičke (Sentinel-2; S2) satelitske snimke visoke rezolucije (URL 1). Razlika između prikupljanja podataka pomoću Sentinel-1 i Sentinel-2 satelita je u uređajima za primanje i registriranje elektromagnetskih valova koji mogu biti aktivni i pasivni. Aktivni sustavi sami proizvode zračenje, zrake odašilju u prostor, one se odbijaju od objekata i tako odbijene ih registrira uređaj. Pasivni uređaji primaju elektromagnetske zrake koje dolaze od objekata, bilo kao njegovo vlastito zračenje (npr. toplinske zrake), bilo kao odbijene zrake sunčevog zračenja (vidljivi spektar i dr). Za klasifikaciju zemljišnog pokrova koriste se optičke snimke (Gašparović i dr., 2018), radarske snimke (Balzter i dr., 2015) te njihova integracija (Clerici i dr., 2017).

U ovom radu usporediti će se Sentinel-1 i Sentinel-2 snimke jednake prostorne rezolucije od 10 m pomoću različitih metoda strojnog učenja. Uz Random Forest (RF) metodu, koja se najčešće koristi za nadziranu klasifikaciju, koristiti će se još Support Vector Machine (SVM) i XGBoost (XGB) metode.

2. PODRUČJE ISTRAŽIVANJA I PODACI

Područje istraživanja ovog rada je šire područje grada Zagreba, koji je ujedno glavni i najveći grad Republike Hrvatske. Prema popisu stanovništva iz 2011. godine, grad broji 790 017 stanovnika. Nalazi se u kontinentalnoj središnjoj Hrvatskoj i smjestio se podno južnih obronaka Medvednice, sa centralnim urbanim dijelom te rijekom Savom i poljoprivrednim površinama na južnom dijelu grada. Za potrebe ovog rada, područje istraživanja iznosi 840 km² (21 km x 40 km), kojim je obuhvaćen manji dio planine Medvednice, a u većem udjelu je pokriven južni nizinski dio (Slika 1).



Slika 1: Prikaz područja istraživanja na: (a) Sentinel-1 radarskom snimku za datum 19.06.2018. ; (b) Sentinel-2 multispektralnom satelitskom snimku za datum 12.06.2018.

Na Sentinel-1 satelitima ugrađen je C-SAR (engl. *Synthetic Aperture RADAR*) koji radi u području valnih duljina od 5,55 cm. Dobiveni podaci mogu biti u različitim kombinacijama horizontalne i vertikalne polarizacije, prostorne rezolucije između 5 i 100 m te veličini sekcije od 20 do 400 km (Rubinić, 2018). Za potrebe ovog istraživanja korištena je GRD snimka (engl. *Ground Range Detected*) razine 1 za datum od 19.06.2018. godine. Preuzeta GRD snimka predobrađena je u programskom paketu Sentinel Application Platform (SNAP) verzije 6.0. U SNAP-u izvršena je kalibracija snimke čime se dobio koeficijent povratnog raspršenja, zatim speckle filtracija pomoću Lee prostornog filtera veličine prozora 5 x 5 piksela te ortorektifikacija pomoću SRTM digitalnog modela terena.

Sentinel-2 satelitska snimka razine 1C preuzeta je sa Copernicus Open Access Hub servisa i pomoću Sen2Cor modula unutar SNAP paketa konvertirana iz razine 1C (refleksija na vrhu atmosfere) u razinu 2A (refleksija pri dnu atmosfere). Datum Sentinel-2 snimka je 12.06.2018. sa 0% prekrivenosti oblacima. Za potrebe ovog istraživanja korišteni su spektralni kanali 2, 3, 4 i 8 (plavi, zeleni, crveni i blisko-infracrveni) koji su jednake prostorne rezolucije od 10 m kao i obrađena Sentinel-1 GRD snimka.

Nakon predobrade radarskih i satelitskih snimaka, uzorci za potrebe nadzirane klasifikacije prikupljali su se u Quantum GIS programu verzije 2.18.27, dok se za klasifikaciju snimaka i provjeru točnosti klasifikacije koristio R programski jezik, verzije 3.6.0 unutar grafičkog sučelja Rstudio, verzije 1.0.153.

3. METODE

Za potrebe nadzirane klasifikacije Sentinel-1 i Sentinel-2 snimka, korištene su Random Forest (RF), Support Vector Machine (SVM) i XGBoost (XGB) metode strojnog učenja. Navedene metode imaju mogućnost prepoznavanja konačnih klasa iz trening uzoraka i primjenu navedenog znanja na nove, neklasificirane podatke (Belgiu i Dragut, 2016).

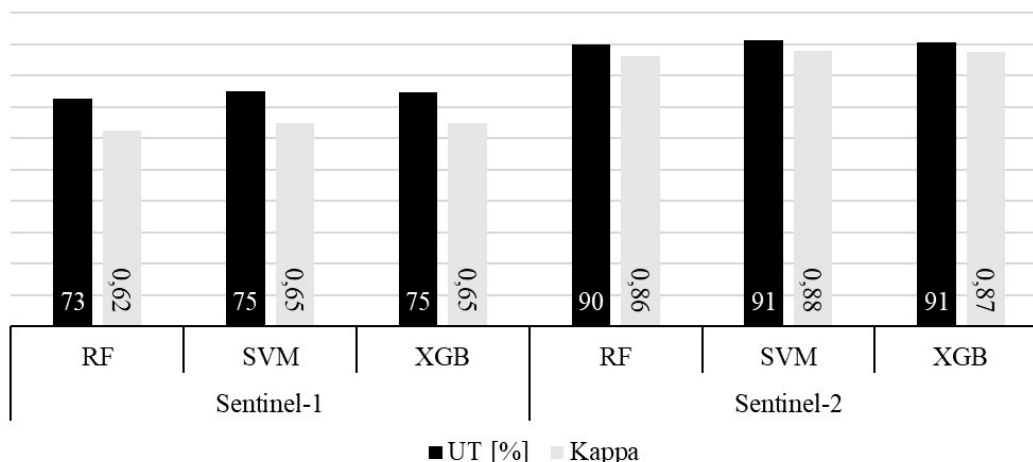
RF metoda temelji se na izgradnji stabla odluke (engl. *decision tree*). Iz trening podataka nasumično se uzimaju uzorci te iz dobivenih nasumičnih uzoraka algoritam izgradi „šumu“ stabala. Konačna klasa u klasifikaciji odabere se na temelju odgovora većine stabala (Breiman, 2001). Prema sličnom istraživanju (Belgiu i Dragut, 2016), broj stabala (*ntree*) odabran je na 500, dok je *mtry* parametar određen kao kvadratni korijen od ukupnog broja ulaznih varijabli. SVM definira hiperravninu koja klasificira sve vektore iz trening skupa podataka u dvije ili više klasa. Budući da granica navedene ravnine između zadanih klasa nije linearna, koristi se pristup jezgre (engl. *kernel*). Osnovna ideja je da se svi unutarnji produkti koji se pojavljuju prilikom računanja zamjene s funkcijom K koja se naziva jezgra. Najčešće korištene jezgre su: linearna, polinomna i radijalna. U ovom istraživanju za SVM metodu korištena je radijalna jezgra sa dodatnim argumentima *cost* (C) i širina jezgre (γ) (Noi i Kappas, 2017). XGB pripada *boosting* metodi strojnog učenja unutar koje se novi modeli stvaraju uzimajući u obzir greške dobivene prethodnim modelima (Chen i Guestrin, 2016). Dakle, za razliku od RF metode koja nasumično bira uzorke, XGB metoda stvara novi model pomoću minimiziranja greške dobivene od prethodnog modela, sve do trenutka u kojem se daljnja poboljšanja ne mogu više dobiti (engl. *gradient descent algorithm*).

U okviru ovog istraživanja, zemljišni pokrov podijeljen je u pet klasa: voda, zemlja, šuma, izgrađeno zemljište i niska vegetacija. Za potrebe nadzirane klasifikacije, sa Sentinel-2

snimka prikupljeno je 600 uzoraka koji su ravnomjerno raspoređeni unutar klasa i dovoljno veliki da zadovolje sve veći broj različitih dimenzija podataka. Uzorci su nasumično odabrani i podijeljeni u omjeru 70% za potrebe učenja modela, naspram 30% uzoraka koji su se koristili za provjeru točnosti klasifikacije. Za provjeru točnosti korištena je matrica konfuzije, unutar koje podaci na dijagonalama predstavljaju ispravno klasificirane piksele, dok su pikseli izvan dijagonale pogrešno klasificirani u drugu klasu. Pored ukupne točnosti klasifikacije, u istraživanju će se prikazati korisničke (retci matrice konfuzije) i proizvodne točnosti (stupci matrice konfuzije) dobivene za pojedinu klasu ovisno o korištenom senzoru za klasifikaciju (Foody, 2002).

4. REZULTATI I DISKUSIJA

Nakon izrade klasifikacije zemljišnog pokrova za Sentinel-1 i Sentinel-2 snimaka, analiza točnosti klasifikacije izvršena je na 200 nasumično odabranih uzoraka. Ukupna točnost klasifikacije definirana je kao omjer ispravno klasificiranih piksela naspram ukupnog broja piksela. Kod Sentinel-1 snimka ukupna točnost (UT) klasifikacije iznosi od 73% do 75% ovisno o korištenoj metodi strojnog učenja, dok je Kappa koeficijent u rasponu od 0,62 do 0,65. Veliko povećanje točnosti nastalo je kod klasifikacije Sentinel-2 snimka - UT je u rasponu od 90% do 91%, a Kappa koeficijent iznosi od 0,86 do 0,88 za RF, SVM i XGB metodu (Slika 2). Ukoliko usporedimo pojedinu metodu korištenu za klasifikaciju, kod Sentinel-1 najviša točnost dobivena je pomoću SVM i XGB metode, a najniža kod RF; kod Sentinel-2 najvišu točnost dobilo se SVM metodom, zatim XGB, i najniža točnost je ostvarena sa RF metodom.



Slika 2: Ukupna točnost (UT) i kappa koeficijent (Kappa) klasifikacije Sentinel-1 i Sentinel-2 snimka dobiveni pomoću različitih metoda strojnog učenja.

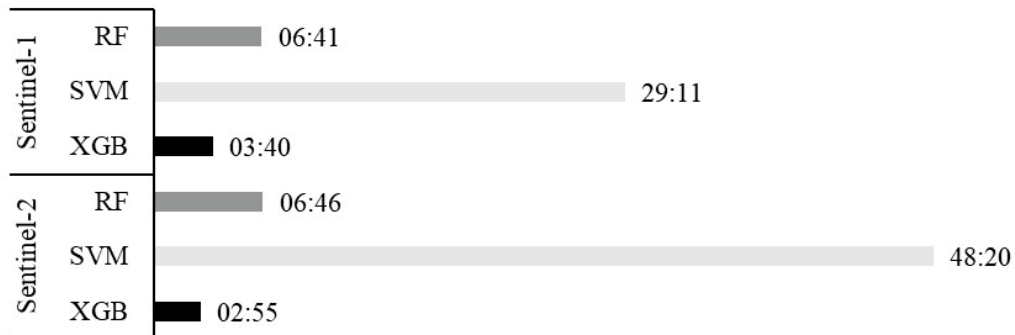
Pregledom korisničke (engl. *user's accuracy*) i proizvodne točnosti (engl. *producer's accuracy*) može se detaljnije proučiti uspješnost klasifikacije prema pojedinim klasama. Pritom se vrijednost korisničke točnosti (KT) dobije dijeljenjem broja ispravno klasificiranih piksela koji pripadaju nekoj klasi sa zbrojem vrijednosti u redovima te iste klase, dok se vrijednost proizvodne točnosti (PT) dobije dijeljenjem broja ispravno klasificiranih piksela koji pripadaju nekoj klasi sa zbrojem vrijednosti u stupcu te iste klase (Story i Congalton, 1986). U nastavku biti će prikazane KT i PT za klasificirani Sentinel-1 i Sentinel-2 snimak pomoću RF, SVM i XGB metode (Tablica 1).

Tablica 1: Korisnička i proizvodna točnost klasifikacije Sentinel-1 i Sentinel-2 snimka pomoću RF, SVM i XGB metode.

Sentinel-1										
Metoda \ Klasa	Voda		Zemlja		Šuma		Izgrađeno		Niska veg.	
	KT	PT	KT	PT	KT	PT	KT	PT	KT	PT
RF	98,6	74,5	49,8	73,9	78,6	84,6	79,1	62,5	53,1	50,1
SVM	98,9	74,1	57,0	72,9	77,1	89,6	89,7	60,7	57,8	53,0
XGB	98,5	74,4	53,9	74,2	78,9	87,2	85,6	62,4	56,9	55,0
Sentinel-2										
Metoda \ Klasa	Voda		Zemlja		Šuma		Izgrađeno		Niska veg.	
	KT	PT	KT	PT	KT	PT	KT	PT	KT	PT
RF	99,8	94,0	86,8	74,8	97,8	93,9	94,5	87,3	79,1	90,8
SVM	99,9	94,6	85,0	87,4	97,7	93,8	97,8	82,8	77,8	92,7
XGB	99,8	95,2	84,5	83,4	98,0	93,7	97,2	84,4	78,2	91,5

Iz Tablice 1, ukoliko usporedimo KT prema pojedinim klasama, na Sentinel-1 sa visokom razinom točnosti klasificirane su voda, šuma i izgrađeno zemljište, dok su klase zemlja i niska vegetacija klasificirane sa nižom točnošću. Pritom je većina neispravno klasificiranih piksela iz klase zemlja prešlo u klasu izgrađeno, dok su pikseli iz niske vegetacije prešli u šumu. Navedeni slučajevi događaju se budući da radari kao aktivni senzori opažaju jačinu povratnog signala. Nasuprot tome, optički senzori registriraju energiju koja dolazi od Sunca do Zemlje, i odbija se u vidljivom i infracrvenom dijelu elektromagnetskog spektra. Upravo zbog blisko-infracrvenog kanala Sentinel-2 je uspješno klasificirao šumska područja za razliku od Sentinel-1 te je sa visokom točnošću klasificirao vodu, zemlju i izgrađeno. Kod Sentinel-2, niska vegetacija uglavnom je pogrešno klasificirana u šumu, i manjim dijelom u klasu zemlja. Kada usporedimo rezultate klasifikacija i pregledom znanstvene literature, Sentinel-1 GRD proizvodi se koriste za kartiranje poplavljenih područja (Tsyganskaya i dr., 2018), dok se Sentinel-2 snimke, između ostalog, koriste za detekciju promjena zemljišnog pokrova kroz vrijeme (Belgiu i Csillik, 2018). U nastavku biti će prikazana vremena koja su bila potrebna u R programskom jeziku za pojedinu metodu strojnog učenja prema pojedinom snimku (Slika 3).

Vrijeme [mm:ss]



Slika 3: Prikaz vremena potrebnom za računanje Sentinel-1 i Sentinel-2 klasifikacije za pojedinu metodu strojnog učenja.

Iz prethodne slike vidljivo je da je za klasifikaciju Sentinel-1 i Sentinel-2 snimka najmanje vremena trebalo XGB metodi, potom RF metodi, dok je najviše vremena bilo potrebno SVM metodi, iako je upravo SVM postizao najviše točnosti. Razlog dužeg vremena koje je potrebno SVM metodi je taj što je navedena metoda prvotno dizajnirana za binarne (*two-class*) probleme (Cortes i Vapnik, 1995). Nasuprot tome, XGB pripada *boosting* tehnici strojnog učenja unutar koje se novi modeli stvaraju uzimajući u obzir greške dobivene prethodnim modelima i samim time se skraćuje vrijeme potrebno za izračun klasifikacije.

5. ZAKLJUČAK

U ovom radu uspoređeni su rezultati klasifikacije zemljišnog pokrova na radarskim (Sentinel 1; S1) i optičkim (Sentinel-2; S2) satelitskim snimkama jednake prostorne rezolucije od 10 m pomoću različitih metoda strojnog učenja. Područje istraživanja je grad Zagreb sa okolinom na kojem je prikupljeno 600 uzoraka za nadziranu klasifikaciju zemljišnog pokrova koji je podijeljen u pet klasa (voda, zemlja, šuma, izgrađeno zemljište i niska vegetacija). Usporedbom prosječne vrijednosti Kappa koeficijenta izračunatim za Sentinel-1 i Sentinel-2 klasifikaciju, može se zaključiti da je za Sentinel-1 postignuta visoka točnost klasifikacije koja iznosi 0,64, dok je za Sentinel-2 postignuta vrlo visoka točnost klasifikacije u vrijednosti od 0,87. Pritom je SVM metoda postigla najvišu točnost, potom XGB metoda te najnižu točnost u ovom istraživanju postigla je RF metoda. Uvidom u točnosti koje su postignute za pojedinu klasu, Sentinel-1 je imao problema sa klasificiranjem niske vegetacije i zemlje, dok je Sentinel-2 uglavnom pogrešno klasificirao nisku vegetaciju. Pored točnosti koje su postignute za klasifikaciju zemljišnog pokrova, veoma je važan i vremenski okvir koji je potreban za izračun pojedine klasifikacije. Prosječna vremena iznose 6 min, 40 min i 3 min za RF, SVM i XGB metodu, tako da je potrebno uzeti u obzir kompromis između točnosti i vremena koje je potrebno za klasifikaciju.

Važno je naglasiti kako je cijelo istraživanje napravljeno pomoću besplatno dostupnih podataka u okviru Copernicus programa te u programima otvorenog koda. U budućnosti, planirano je istražiti vremenske analize Sentinel-1 i Sentinel-2 snimaka, kao i njihovu zajedničku integraciju u svrhu poboljšanja točnosti klasifikacije prema pojedinim klasama.

ZAHVALA

Ovo istraživanje provedeno je u sklopu istraživačkog projekta GEMINI (Geoprostorno praćenje zelene infrastrukture na temelju terestričkih, zračnih i satelitskih snimaka, IP-2016-06-5621) što ga financira Hrvatska zaklada za znanost.

LITERATURA

- Balzter, H., Cole, B., Thiel, C., Schmullius, C. (2015): Mapping CORINE land cover from Sentinel-1A SAR and SRTM digital elevation model data using random forests. *Remote Sensing*, 7, 11, 14876-14898.
- Belgiu, M., Drăguț, L. (2016): Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24.
- Belgiu, M., Csillik, O. (2018): Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, 204, 509-523.
- Breiman, L. (2001): Random forests. *Machine learning*, 45, 1, 5-32.
- Chen, T., Guestrin, C. (2016): Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, ACM, 785-794.
- Clerici, N., Valbuena Calderón, C. A., Posada, J. M. (2017): Fusion of Sentinel-1A and Sentinel-2A data for land cover mapping: a case study in the lower Magdalena region, Colombia. *Journal of Maps*, 13, 2, 718-726.
- Cortes, C., Vapnik, V. (1995): Support-vector networks. *Machine learning*, 20, 3, 273.
- Foody, G. M. (2002): Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, 1, 185-201.
- Gašparović, Mateo; Dobrinić, Dino; Medak, Damir (2018): Urban vegetation detection based on the land-cover classification of PlanetScope, RapidEye and Worldview-2 Satellite Imagery, *International Multidisciplinary Scientific Geoconference SGEM 2017*, 18, 2.3, 249-256.
- Mulla, D. J. (2013): Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems engineering*, 114, 4, 358-371.
- Rubinić, A. (2018): Kartiranje poplave na Grobničkom polju u studenom 2016. na temelju Sentinel-1 satelitskog snimka. *Hrvatske vode*, 26, 104, 129-134.
- Story, M., Congalton, R. G. (1986): Accuracy assessment: a user's perspective. *Photogrammetric Engineering and remote sensing*, 52, 3, 397-399.
- Thanh Noi, P., Kappas, M. (2018): Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, 18, 1, 18.
- Tsyganskaya, V., Martinis, S., Marzahn, P., Ludwig, R. (2018): SAR-based detection of flooded vegetation—a review of characteristics and approaches. *International journal of remote sensing*, 39, 8, 2255-2293.
- Valozić, Luka (2015): Objektno orijentirana klasifikacija zemljišnoga pokrova pomoću multispektralnih satelitskih snimaka – primjer Grada Zagreba, *Hrvatski geografski glasnik*, 76, 2, 27-38.
- URL 1: Copernicus in Brief, <https://www.copernicus.eu/en/about-copernicus/copernicus-brief>

Land cover classification using data from Sentinel satellite missions

ABSTRACT: *Main goal of this research is land cover classification using data acquired from Sentinel-1 (Sentinel-1; S1) and Sentinel-2 (Sentinel-2; S2) satellites. Copernicus is the European Union's Earth Observation Programme to develop European information service based on Earth observation satellites, in situ data and from other contributing missions. Within the Programme, users can obtain full, free and open data from radar (S1) and optical (S2) satellite imagery of high resolution. Radar provides it's own energy source for illumination. Therefore measurements can be obtained anytime, regardless of the time of day or season, and with longer wavelenghts it is practically independent from weather conditions. Sentinel-1satellites were launched in 2014 and 2016; every satellite carry an active sensor (Synthetic Aperture RADAR – SAR) and imagery is available every six days. Sentinel-2is a multispectral high resolution satellite mission (comprised of S2-A and S2-B satellite) that samples Earth in 13 spectral bands with spatial resolution of 10 m, 20 m and 60 m.*

In this research comparison of radar and optical satellite imagery with spatial resolution of 10 m using different machine learning methods is investigated. Along with Random Forest (RF), XGBoost and Support Vector (SVM) will be tested. RF and XGBoost combine predictions from multiple learning algorithms (ensemble learning), while SVM identifies the optimal hyper-plane which segregates vectors from training data set into two or more land cover classes. Main goal of this research is to compare land cover classification results from radar and optical satellite imagery and to define their applications for future research.

Ključne riječi: classification, Random Forest, Sentinel, Support Vector Machine, XGBoost.