

Univerza
v Ljubljani

Fakulteta za
*gradbeništvo in
geodezijo*



INTERDISCIPLINARNI
DOKTORSKI ŠTUDIJSKI
PROGRAM VARSTVO OKOLJA

Kandidat:

PETER LAMOVEC, univ. dipl. inž. geod.

**DOLOČITEV POPLAVLJENIH OBMOČIJ Z UPORABO
DALJINSKEGA ZAZNAVANJA
IN STROJNEGA UČENJA**

Doktorska disertacija števil: 1/VO

**DETERMINATION OF FLOODED AREAS USING
REMOTE SENSING AND MACHINE LEARNING**

Doctoral thesis No.: 1/VO

Soglasje k temi doktorske disertacije je dala Komisija za doktorski študij na 20. redni seji, 21. septembra 2011. Za mentorja je bil imenovan izr. prof. dr. Krištof Oštir, CO Vesolje-SI, ZRC SAZU in UL FGG, za somentorja prof. dr. Matjaž Mikoš, UL FGG.

Ljubljana, 16. julij 2013

Univerza
v Ljubljani

Fakulteta za
*gradbeništvo in
geodezijo*



Komisijo za oceno ustreznosti teme doktorske disertacije v sestavi:

- izr. prof. dr. Krištof Oštir, CO Vesolje-SI, UL FGG in ZRC SAZU,
- prof. dr. Matjaž Mikoš, UL FGG,
- doc. dr. David Hladnik, UL BF,
- prof. dr. Jože Rakovec, UL FMF,

je imenoval Senat Fakultete za gradbeništvo in geodezijo na 20. redni seji, 20. aprila 2011.

Poročevalce za oceno doktorske disertacije v sestavi:

- doc. dr. Mojca Kosmatin Fras, UL FGG,
- doc. dr. David Hladnik, UL BF,
- prof. dr. Jože Rakovec, UL FMF,

je imenoval Senat Fakultete za gradbeništvo in geodezijo na 40. redni seji, 24. aprila 2013.

Komisijo za zagovor doktorske disertacije v sestavi:

- prof. dr. Matjaž Mikoš, dekan UL FGG, predsednik in somentor,
- izr. prof. dr. Krištof Oštir, CO Vesolje-SI, ZRC SAZU, UL FGG, somentor,
- doc. dr. Mojca Kosmatin Fras, UL FGG,
- doc. dr. David Hladnik, UL BF,
- prof. dr. Jože Rakovec, UL FMF,

je imenoval Senat Fakultete za gradbeništvo in geodezijo na 2. redni seji, 26. junija 2013.

Univerza
v Ljubljani

Fakulteta za
*gradbeništvo in
geodezijo*



IZJAVA O AVTORSTVU

Podpisani **Peter Lamovec**, univ. dipl. inž. geod., izjavljam, da sem avtor doktorske disertacije z naslovom: **Določitev poplavljenih območij z uporabo daljinskega zaznavanja in strojnega učenja.**

Izjavljam, da je elektronska različica v vsem enaka tiskani različici.

Izjavljam, da dovoljujem objavo elektronske različice v repozitoriju UL FGG.

Ljubljana, 16. julij 2013

.....
(podpis)

ERRATA

Stran z napako	Vrstica z napako	Namesto	Naj bo
----------------	------------------	---------	--------

BIBLIOGRAFSKO-DOKUMENTACIJSKA STRAN IN IZVLEČEK

UDK:	519.682:528.8:556.166(043.3)
Avtor:	Peter Lamovec, univ. dipl. inž. geod.
Mentor:	izr. prof. dr. Krištof Oštir
Somentor:	prof. dr. Matjaž Mikoš
Naslov:	Določitev poplavljenih območij z uporabo daljinskega zaznavanja in strojnega učenja
Obseg in oprema:	119 str., 17 pregl., 25 sl., 6 pril.
Ključne besede:	strojno učenje, odločitveno drevo, klasifikacija, satelitski posnetki, hidrologija, poplave

Izvleček

V disertaciji smo razvili postopek prepoznavanja poplavljenih površin s pomočjo podatkov daljinskega zaznavanja in algoritmov strojnega učenja. Glavni podatki daljinskega zaznavanja, uporabljeni v doktorski disertaciji, so bili optični posnetki satelitskih sistemov SPOT in RapidEye, poleg njih pa sta bila uporabljena še digitalni model višin in mreža vodotokov. Algoritmi strojnega učenja omogočajo določitev najpomembnejših atributov za klasifikacijo poplavljenih površin in njihovih mejnih vrednosti, ki določajo, ali bo območje uvrščeno v razred poplavljenih ali nepoplavljenih površin. Pri tem se srečujemo z več vprašanji, ki so povezana s pripravo potrebnih podatkov, sestavo ustreznega učnega vzorca in izbiro pravilnega algoritma. Vsako od teh vprašanj zahteva veliko časa, če nanje nismo pripravljeni oz. nanje ne znamo odgovoriti. Čas in hiter odziv pa sta ob nastanku poplav ključnega pomena. Doktorska disertacija bralca seznani z glavnimi značilnostmi hudourniškega oz. kraškega tipa poplav. Popelje ga skozi več različnih klasifikacij poplavljenih površin, ki se razlikujejo po uporabljenih podatkih, učnih vzorcih in algoritmi, in mu predstavi probleme, ki se pri tem pojavljajo. Klasifikacije hudourniškega tipa poplav so predstavljene na primeru poplav v Železnikih in klasifikacije kraškega tipa poplav na primeru poplav na Ljubljanskem barju. Rezultat je določitev najuspešnejše klasifikacije za hitro in natančno določitev poplavljenih površin na obeh študijskih območjih.

BIBLIOGRAPHIC-DOCUMENTAL INFORMATION AND ABSTRACT

UDC: 519.682:528.8:556.166(043.3)
Author: Peter Lamovec
Supervisor: Assoc. Prof. Krištof Oštir, Ph. D.
Co-supervisor: Prof. Dr. Matjaž Mikoš, Ph. D.
Title: Determination of flooded areas using remote sensing and machine learning
Notes: 119 p., 17 tab., 25 fig., 6 app.
Key words: machine learning, decision tree, classification, satellite images, hydrology, floods

Abstract

The thesis presents a procedure for determining flooded areas using remote sensing data and machine learning algorithms. Main used remote sensing data were optical images of SPOT and RapidEye satellite systems. Other used data were digital elevation model and hydrological network. The classification is directed by attributes and their threshold values, which define the boundary between flooded and unflooded classes. The most influential attributes are defined by machine learning algorithms. During the process of determining flooded areas many questions are happening about used applicable data, suitable training data and selection of correct algorithms. While quick response is crucial when flood happens, a lot of time can be spent when the answers are unknown. So, the thesis presents the most important characteristics of two typed of floods: flash- and karst floods, the reader is acquainted with different classifications of flooded areas, distinguishing in used data, training samples and algorithms and the difficulties which are happening on both type of floods are described. Classifications of flash floods are presented on the case study of Železniki (2007) and classifications of karst floods are presented on the case study of Ljubljana Moor. The result is determination of the procedure for quick and accurate classification of flooded areas on both study areas.

ZAHVALA

Izdelava doktorske disertacije je dolga in zahtevna naloga, ki ne bi bila mogoča brez pomoči številnih posameznikov in različnih ustanov, ki so mi na tej poti pomagali. Posebej bi se rad zahvalil mentorju dr. Krištofu Oštirju in somentorju dr. Matjažu Mikošu, ki sta me s svojo strokovnostjo in hkrati preprostostjo navdušila za izbrano raziskovalno področje in me na tej poti vseskozi pravilno usmerjala. Zahvalil bi se tudi vsem sodelavcem na Inštitutu za antropološke in prostorske študije, ki so mi pomagali pri reševanju najrazličnejših problemov preko celotnega obdobja in v družbi katerih je bilo delo vedno zanimivo. Zahvaljujem se tudi dr. Sašu Džeroskemu in dr. Dragiju Kocevu iz Inštituta Jožef Štefan za pomoč pri razumevanju strojnega učenja. Hvala pa tudi vsem drugim, ki jim ni bilo škoda časa, ki so ga porabili za potrpežljivo odgovarjanje na vprašanja, pa naj je šlo za elektronsko pošto, telefonske pogovore ali za srečanja na različnih sestanku. Delček prav vsakega izmed njih se skriva tudi v tej doktorski disertaciji.

Na koncu bi se rad zahvalil še vsem svojim domačim in tistim, ki jih imam najraje. Hvala, da ste verjeli vame, me spodbujali in skrbeli, da so bili tudi najnapornejši trenutki lepi.

KAZALO

1	UVOD	1
1.1	Ozadje problema.....	2
1.2	Raziskovalne hipoteze in cilji disertacije	3
1.2.1	Raziskovalne hipoteze in njihova utemeljitev	3
1.2.2	Cilji	5
1.3	Zgradba doktorske disertacije.....	6
2	UPORABA METOD ZA PREPOZNAVANJE POPLAVLJENIH POVRŠIN	7
2.1	Daljinsko zaznavanje.....	7
2.1.1	Podatki daljinskega zaznavanja	8
2.1.2	Pregled stanja na področju prepoznavanja poplavljenih površin z uporabo podatkov daljinskega zaznavanja	12
2.2	Strojno učenje	17
3	UPORABA STROJNEGA UČENJA.....	20
3.1	Postopek klasifikacije z uporabo strojnega učenja	21
3.1.1	Določitev vzorca s slučajno generiranimi točkami/segmenti	22
3.1.2	Določitev atributov za učenje	22
3.1.3	Razdelitev vzorca v množico za učenje in testno množico	26
3.1.4	Izbira primernega klasifikacijskega algoritma in izvedba učenja.....	26
3.1.5	Ocena natančnosti modela na testnih podatkih.....	27
3.1.6	Klasifikacija z uporabo izbranega klasifikacijskega modela.....	27
3.2	Algoritmi strojnega učenja za klasifikacijo poplavljenih površin	28
4	ŠTUDIJSKO OBMOČJE	33
5	KARTIRANJE POPLAV	35
5.1	Železniki - primer hudourniških poplav	36
5.1.1	Podatki	36
5.1.2	Metode dela	40
5.1.3	Rezultati.....	47
5.1.4	Razprava	56
5.2	Ljubljansko barje - primer kraškega tipa poplav	60
5.2.1	Podatki	60
5.2.2	Metode dela	62
5.2.3	Rezultati.....	68
5.2.4	Razprava	75
6	MOŽNOSTI UPORABE REZULTATOV IN NADALJNJE DELO	80
7	ZAKLJUČEK.....	83
8	POVZETEK	87
9	SUMMARY	89
10	VIRI.....	91
11	PRILOGE	100

KAZALO PREGLEDNIC

Preglednica 1: Izsek preglednice z atributi učnega vzorca za prepoznavanje poplavljenih površin ob poplavah v Železnikih 2007. Preglednica prikazuje deset od 255 vrstic/učnih primerov, vključenih v proces učenja.	23
Preglednica 2: Algoritmi strojnega učenja, uporabljeni za klasifikacijo poplavljenih površin na območju Železnikov in Ljubljanskega barja. Zadnja dva stolpca označujeta obliko učnega vzorca, ki ga predstavljajo točke oz. segmenti.	30
Preglednica 3: Pridobljeni satelitski posnetki za opazovanje poplavljenih površin na območju Železnikov	37
Preglednica 4: Podatkovni sloji in njihovi atributi, uporabljeni za prepoznavanje poplavljenih površin na območju Železnikov. Pri posameznem kanalu posnetka SPOT 5 so v oklepaju podana spektralna območja zaznanega elektromagnetnega valovanja v nanometrih (nm).	39
Preglednica 5: Strojno učenje na območju Železnikov. Preglednica prikazuje različne kombinacije strojnega učenja, ki so bile preizkušene glede na ločljivosti uporabljenega DMV, gostoto učnega vzorca, obliko učnih primerov in vrsto uporabljenih algoritmov.	42
Preglednica 6: Uspešnost določitve poplav z uporabo različnih metod strojnega učenja. Med metodami za kombiniranje klasifikatorjev so prikazani rezultati z 10 drevesi.	51
Preglednica 7: Matrika napak za oceno natančnosti klasifikacije na območju Železnikov.	51
Preglednica 8: Natančnost določitve poplavljenih površin z uporabo metode naključnih gozdov (random forest) in različnim številom dreves.	52
Preglednica 9: Prikaz atributov, ki sodelujejo pri klasifikacijah poplavljenih površin na območju Železnikov z učenjem na učnem vzorcu točk in različnimi algoritmi. Vrednosti v preglednici predstavljajo število modelov, v katerih se pojavi posamezen atribut. Večje število pojavljanj pomeni večji vpliv atributa na klasifikacijo.	57
Preglednica 10: Uporabljeni satelitski in letalski posnetki za opazovanje poplavljenih površin na območju Ljubljanskega barja 2010.	61
Preglednica 11: Podatkovni sloji in njihovi atributi, uporabljeni za prepoznavanje poplavljenih površin na območju Ljubljanskega barja. Pri posameznem kanalu posnetka RapidEye so v oklepaju podana tudi spektralna območja zaznanega elektromagnetnega valovanja v nanometrih (nm).	62
Preglednica 12: Strojno učenje na območju Ljubljanskega barja. Preglednica prikazuje različne kombinacije strojnega učenja, ki so bile preizkušene glede na obliko učnih primerov in vrsto uporabljenih algoritmov.	62
Preglednica 13: Uspešnost določitve poplav z uporabo različnih algoritmov strojnega učenja, ki je potekalo na učnem vzorcu, sestavljenem iz segmentov.	70
Preglednica 14: Uspešnost določitve poplav z uporabo različnih algoritmov strojnega učenja, ki je potekalo na učnem vzorcu, sestavljenem iz točk.	70
Preglednica 15: Matrika napak za oceno natančnosti klasifikacije na območju Ljubljanskega barja.	71
Preglednica 16: Časovna vrsta uporabljenih radarskih satelitskih posnetkov.	74
Preglednica 17: Prikaz atributov, ki sodelujejo pri klasifikacijah poplavljenih površin na območju Ljubljanskega barja z učenjem na učnem vzorcu točk in različnimi algoritmi. Vrednosti v preglednici predstavljajo število modelov, v katerih se pojavi posamezen atribut. Večje število pojavljanj pomeni večji vpliv atributa na klasifikacijo.	76

LIST OF TABLES

Table 1: Part of the table with attributes of training samples for determination of flooded areas in Železniki 2007. Table presents 10 out of 255 rows/training samples included in the learning process.	23
Table 2: Machine learning algorithms used for the classification of flooded areas in the area of Železniki and Ljubljana Moor. Two different samples of points and segments were used (signed in the last two columns).	30
Table 3: Acquired satellite images for determination of flooded areas in the area of Železniki.....	37
Table 4: Data and their attributes used for determination of flooded areas in the area of Železniki. Wavelength of each spectral band of SPOT 5 sensors is given in brackets in the second column of the table.	39
Table 5: Machine learning in study area of Železniki. Table presents different combinations of DEM, training samples and machine learning algorithms which were used for machine learning process.	42
Table 6: Accuracy of determination of flooded areas using different machine learning methods. Ensembles with ten trees are shown in the table.....	51
Table 7: Confusion matrix to define classification accuracy in the study area of Železniki.....	51
Table 8: Accuracy of determination of flooded areas using random forest method with different number of trees.....	52
Table 9: Attributes included into the model for classification of flooded areas in the area of Železniki. The learning process was performed using sample of points and different algorithms. Values in the table present the number of models where the attributes are taking part. Larger values of the attributes show higher impact on the classification process.	57
Table 10: A list of satellite images that were used for observation of the floods in the Ljubljana Moor area.....	61
Table 11: Data and their attributes used for determination of flooded areas in the area of Ljubljana Moor. Wavelengths of each spectral band of RapidEye sensors are given in brackets in the second column. ..	62
Table 12: Machine learning in study area of Ljubljana Moor. Table presents different combinations training samples and machine learning algorithms which were used for machine learning process.	62
Table 13: Accuracy of determination of flooded areas using different machine learning methods with the training set of segments.	70
Table 14: Accuracy of determination of flooded areas using different machine learning methods with the training set of points.	70
Table 15: Confusion matrix to define classification accuracy in the study area of Ljubljana moor.	71
Table 16: List of time series of the radar satellite images to observe flooded areas in time.....	74
Table 17: Attributes included into the model for classification of flooded areas in the area of Ljubljana Moor. The learning process was performed using sample of points and different algorithms. Values in the table present the number of models where the attributes were taking part. Larger values of the attributes show higher impact on the classification process.	76

KAZALO SLIK

Slika 1: Shematski prikaz metod podatkovnega rudarjenja. Zadnja vrstica prikazuje pet primerov postopka klasifikacije: z nevronskimi mrežami, Bayesovimi mrežami, odločitvenimi drevesi, metodo podpornih vektorjev in učenjem na osnovi primerov (Maimon in Rokach, 2005).	29
Slika 2: Lokaciji študijskih območij: zgornji del doline Selške Sore z Železniki in Ljubljansko barje z Ljubljano. Zgoraj: Prikaz na državnem ozemlju Slovenije. Spodaj: Prikaz na državni topografski karti DTK 50 (GURS, 2005).	33
Slika 3: Shematski prikaz postopka prepoznavanja poplavljenih površin od priprave podatkov do izdelave karte.....	36
Slika 4: Študijsko območje Železnikov na orto foto posnetku (GURS, 2006). Zgoraj: Celotno obravnavano območje zgornjega dela Selške doline, ki ga pokriva DMV ločljivosti 12,5 m. Spodaj: Izsek manjšega območja, za katerega je na voljo tudi DMV višje ločljivosti 0,5 m.	40
Slika 5: Učni vzorec z 255 točkami za učenje, razporejenih po celotnem obravnavanem območju na barvnem RGB večspektralnem posnetku SPOT, prikazanim s kombinacijo kanalov 4, 3 in 2. Rdeče točke (24) ležijo na poplavljenih in rumene (231) na nepoplavljenih tleh.	45
Slika 6: Učni vzorec s 145 točkami za učenje, razporejenih na manjšem obravnavanem območju. Rdeče točke (50) ležijo na poplavljenih in rumene (95) na nepoplavljenih tleh.	45
Slika 7: Učni vzorec s segmenti. Učni vzorec sestavlja 255 segmentov, razporejenih po celotnem obravnavanem območju, in 143 segmentov, razporejenih na manjšem izseku (območje v okvirju).	46
Slika 8: Odločitveni drevesi, pridobljeni z učenjem na digitalnem modelu višin DMV 0,5 (levo) in DMV 12,5 (desno). V obeh primerih je učenje potekalo z uporabo algoritma J48 in 145 učnih točk.	48
Slika 9: Odločitveni drevesi, pridobljeni z učenjem na učnem vzorcu segmentov (levo) oz. točk (desno). V obeh primerih je učenje potekalo na DMV 0,5 z uporabo algoritma J48.	50
Slika 10: Karta poplavljenih površin na območju Železnikov, pridobljena z algoritmom J48, 145 učnimi točkami in DMV 0,5. Poplavljene površine so predstavljene na večspektralnem satelitskem posnetku SPOT, prikazanim s kombinacijo kanalov kratkovalovne IR, bližnje IR in rdeče svetlobe.	53
Slika 11: Karta poplavljenih površin celotnega obravnavanega območja, pridobljena z algoritmom J48 (zgoraj). Spodaj sta prikazana odločitveno drevo in uporabljeni učni vzorec s 145 točkami, razporejenimi na manjšem izseku.	55
Slika 12: Spektralni odboj na različnih pokrovnostih tal (voda, zelena travnik, suh travnik, njiva) glede na valovno dolžino EMV (Ashraf in sod., 2011).	58
Slika 13: Spektralni odboj na vegetaciji. Lepo se vidi vpliv klorofila, celične strukture in vsebnosti vode v listih na odbojnost v različnih območjih vidne in IR svetlobe (Ashraf in sod., 2011).	58
Slika 14: Poplavljene površine pod gozdom. S puščico so označena mesta uspešno prepoznanega robu poplavljenega območja na mestih prekritih z gozdom (prikaz na večspektralnem satelitskem posnetku SPOT - levo oz. letalskem posnetku DOF - desno).	59
Slika 15: Poplavljene površine na pozidanih območjih. Manjših stavb in poplavljenih površin med njimi ni mogoče razlikovati (rdeča puščica). Druge/večje stavbe se enostavno ločijo od poplavljenih površin. .	59
Slika 16: Poplavljene površine na kmetijskih tleh. Do zamenjav kmetijskih površin s poplavljenimi prihaja predvsem na območjih njiv in polj, s katerih je bil pridelek že pospravljen (območja z orno zemljo). ..	60

Slika 17: Vzorec z 201 učno točko, ki so razporejene na obravnavanem območju Ljubljanskega barja na barvnem RGB posnetku RapidEye, prikazanim s kombinacijo kanalov 3, 2, 1. Modre točke (34) ležijo na poplavljenih in zelene (167) na nepoplavljenih tleh.	64
Slika 18: Vzorec z 201 učnim segmentom, ki so razporejeni na obravnavanem območju Ljubljanskega barja na barvnem RGB posnetku RapidEye, prikazanim s kombinacijo kanalov 3, 2, 1. Modro obarvani segmenti (34) ležijo na poplavljenih in zeleni (167) na nepoplavljenih tleh.	65
Slika 19: Radiometrična nehomogenost poplavljenih površin na barvnem satelitskem posnetku RapidEye. Poleg značilne modrozeleno barve so poplavljen površine obarvane tudi s temnejšimi rjavimi in črnimi barvnimi odtenki.	66
Slika 20: Klasifikacija poplavljenih površin z uporabo algoritma J48 na učnem vzorcu točk (zgoraj) in na učnem vzorcu segmentov (spodaj). Na desni strani sta prikazana modela v obliki odločitvenega drevesa.	69
Slika 21: Kartiranje poplavljenih površin z uporabo različnih algoritmov strojnega učenja na učnem vzorcu sestavljenem iz točk.	72
Slika 22: Prikaz umikanja vod s prekrivanjem štirih radarskih satelitskih posnetkov: enega posnetka Envisat in treh posnetkov Radarsat. Prikaz na državni topografski karti DTK 50 (GURS, 2005).	74
Slika 23: Prepoznavanje poplavljenih površin med objekti, ki segajo nad gladino vode. Poplavljen površine so označene z modro barvo (slika zgoraj). Nepoplavljen del predstavljajo višji objekti, ki segajo nad gladino vode, npr. koruznega polja in drevesni nasad (fotografiji spodaj).	77
Slika 24: Poplavljen površine vidne na radarskem posnetku Radarsat. Poplavljen območje ni strnjeno zaradi vmesnih polj in dreves. Vidna je tudi zrnatost – posejanost posnetka s črno-belimi drobnimi strukturami (»posejanost s poprom in soljo«).	78
Slika 25: Karta rabe na območju Ljubljanskega barja. Namen karte je lažja interpretacija kart poplavljenih površin in prepoznavanje težav, ki preprečujejo bolj strnjeno zaznavo poplavljenih površin.	79

LIST OF FIGURES

Figure 1: Scheme of data mining methods. Last row shows five examples of classifications: neural networks, Bayesian networks, decision trees, support vector machines, instance based learning (Maimon in Rokach, 2005).	29
Figure 2: The study area location: the upper Selška Sora River valley with Železniki and Ljubljana Moor with Ljubljana. Above: Location in the area of Slovenia. Below: Presentation on digital topographic map DTK 50 (GURS, 2005).....	33
Figure 3: The process of flooded areas detection from data preparation to map producing.	36
Figure 4: The study area of Železniki on orto foto (GURS, 2006). Above: The whole treated area of the upper Selška Sora River valley covered by digital elevation model (DEM) with resolution 12.5 m. Below: Smaller part of the Selška Sora River valley, where DEM 0,5 exists.	40
Figure 5: Sample of 255 training points lying on the whole treated area on multispectral image SPOT, shown with combination of 4, 3, 2 image channels. Red points (24) are lying in flooded and yellow points (231) in non-flooded areas.....	45
Figure 6: Sample of 145 training points lying in the smaller part of the treated area. Red points (50) are flooded and yellow points (95) are lying in non-flooded areas.	45
Figure 7: Sample of 255 training segments lying on the whole treated area and 143 segments lying on the smaller part (framed area).	46
Figure 8: Decision trees built by 145 sampling points and algorithm J48. DEM 0,5 was used on the left and DEM 12,5 on the right example.	48
Figure 9: Decision trees produced by training sample of segments (left) and training sample of points (right). The DEM 0,5 and algorithm J48 were used in both examples.	50
Figure 10: Map of flooded areas in Železniki produced by algorithm J48, 145 sampling points and DEM 0,5. Flooded areas are represented on multispectral satellite image SPOT in combination of SWIR, NIR and red channels.	53
Figure 11: Map of flooded areas in the whole treated area produced by algorithm J48 (above). Decision tree and sample of 145 training points arranged on the smaller part are shown below.	55
Figure 12: Reflectance spectrum of different land cover types (water, green grass, dry grass, soil) at different wavelengths (Ashraf in sod., 2011).	58
Figure 13: Reflectance spectrum of vegetation. Influence of chlorophyll, cell structure and water content on the reflectance is seen (Ashraf in sod., 2011).	58
Figure 14: Flooding under the forest. Red arrows sign the places where the borders of the flooded areas were detected successfully under the forest (presented on the multispectral SPOT image - left and digital orthophoto DOF - right).	59
Figure 15: Flooding in urban areas. Smaller buildings and flooded areas could not be distinguished completely. Other/larger buildings are distinguished easily from flooded areas.	59
Figure 16: Flooding in agricultural areas. Conflicts between agricultural and flooded areas are happening. Especially cultivated areas of fields are often misclassified as flooded.	60
Figure 17: Sample of 201 training points lying on the treated area of Ljubljana Moor on multispectral image RapidEye, shown with combination of 3, 2, 1 image channels. Blue points (34) are lying on flooded and green points (167) on unflooded areas.	64

Figure 18: Sample of 201 training segments lying on the treated area of Ljubljana Moor on multispectral image RapidEye, shown with combination of 3, 2, 1 image channels. Blue segments (34) are lying on flooded and green segments (167) on unflooded areas.	65
Figure 19: Spectral inhomogeneity of flooded areas on multispectral RapidEye image. Beside usual blue and green colours, flooded areas are coloured also in brown and black.	66
Figure 20: Classification of flooded areas using algorithm J48 with training sample of points (above) and with training sample of segments (below). Decision tree of each classification is shown on the right. ..	69
Figure 21: Maps of flooded areas produced by different machine learning algorithms and training sample of points.	72
Figure 22: Overlaying of four radar images indicates how the floodwater was receding from affected areas. Presentation on digital topographic map DTK 50 (GURS, 2005).	74
Figure 23: Flooded areas between high objects presented with blue colour (above). Non-flooded areas are higher objects extending above water surface, e.g. corn fields, trees (photos below).	77
Figure 24: Flooded areas shown on Radarsat image. Flooded areas are not homogeneous because of fields and trees. The image is affected by speckle noise.	78
Figure 25: Map of land use in the area of Ljubljana Moor to make interpretation of flood maps easier.	79

KAZALO PRILOG

PRILOGA A: Atributi, uporabljeni za strojno učenje, na območju Železnikov in Ljubljanskega barja

PRILOGA B: Množica odločitvenih dreves, pridobljenih z metodo gozdov z 10 drevesi na območju Železnikov

PRILOGA C: Množica odločitvenih dreves, pridobljenih z metodo naključnih gozdov z 10 drevesi na območju Železnikov

PRILOGA D: Množica odločitvenih dreves, pridobljenih z metodo gozdov z 10 drevesi na območju Ljubljanskega barja

PRILOGA E: Množica odločitvenih dreves, pridobljenih z metodo naključnih gozdov z 10 drevesi na območju Ljubljanskega barja

PRILOGA F: Karti poplavljenih površin na območju Železnikov in na območju Ljubljanskega barja

LIST OF APPENDICES

APPENDIX A: Attributes, used in machine learning processes, in the area of Železniki and Ljubljana Moor

APPENDIX B: Ensemble of ten decision trees produced by bagging: study area of Železniki

APPENDIX C: Ensemble of ten decision trees produced by random forest: study area of Železniki

APPENDIX D: Ensemble of ten decision trees produced by bagging: study area of Ljubljana Moor

APPENDIX E: Ensemble of ten decision trees produced by random forest: study area of Ljubljana Moor

APPENDIX F: Maps of flooded areas in the study area of Železniki and Ljubljana Moor.

1 UVOD

Število evidentiranih naravnih nesreč na Zemlji v zadnjem desetletju presega 300 ali celo 400 na leto, ki v povprečju prizadenejo preko 230 milijonov ljudi in zahtevajo več kot sto tisoč smrtnih žrtev letno. V obdobju med letoma 2001 in 2010 so se med vsemi naravnimi nesrečami najpogostejše pojavljale poplave, ki so prizadele tudi največje število svetovnega prebivalstva. Med posameznimi dogodki po velikosti ekonomske škode na prvem mestu izstopajo potresi v razvitih delih sveta. Med naravne nesreče z najhujšimi posledicami za ljudi se uvrščajo še nevihte v obliki tornadov ter tropskih ciklonov in suše (UNISDR, 2011).

Na naravne nesreče moramo gledati tudi z vidika povsem običajnih geofizikalnih pojavov, ki se v naravi pojavljajo že vse od nastanka Zemlje dalje in so oz. še vedno pomembno vplivajo na oblikovanje našega življenjskega okolja (Komac in sod., 2010). Pred več milijoni let so se njihove posledice odražale samo na flori in favni, danes tovrstne dogodke v naravi povezujemo predvsem z njihovimi vplivi na človeka in njegovo imetje (Ayala, 2002). Človek si je za bivanje najprej izbral tista območja, ki so mu nudila najboljše pogoje za preživetje. Tako je med drugim poselil tudi območja v bližini rek, bogata z zadostno zalogo vode in rodovitno prstjo. Njegov način bivanja je bil na začetku prilagojen razmeram v naravi. V naravno okolje je postopoma vpeljal nove, večje in zahtevnejše sisteme, s katerimi je postalo njegovo življenje udobnejše, hkrati pa tudi manj prilagodljivo na nenadne spremembe v okolju. Posledice, ki nastanejo ob pojavu bolj intenzivnega naravnega dogodka, npr. močnih padavinah in poplavih, so zato danes zaradi stanovanjske gradnje, industrijskih con, prometne, energetske, telekomunikacijske infrastrukture, urejenih kmetijskih površin ipd. bistveno hujše, kot so bile nekoč (Skidmore in Toya, 2002). Škodo, ki nastane ob takšnih izrednih dogodkih na s strani človeka urejeni površini, pa moramo razumeti kot posledico obojega: naravnega pojava samega in ranljivosti človeškega sistema (Westgate in O'Keefe, 1976). Izbira prostora namenjenega različnim rabam mora biti zato skrbno premišljena, izvedeni pa morajo biti tudi potrebni varstveni ukrepi v primeru prevelikega tveganja (Mikoš, 2007).

Doktorska disertacija se ukvarja s poplavami, ki so ena izmed najbolj uničujočih naravnih nesreč. Poplave za sabo pustijo veliko opustošenje zemeljskega površja, njihove hude posledice pa se kažejo tako v človeških žrtvah kot v ekonomskih kazalcih na splošno. Med desetimi največjimi svetovnimi naravnimi nesrečami glede na število smrtnih žrtev v letu 2011 je bilo kar šest poplavnih dogodkov, ki so skupaj zahtevali skoraj 3000 življenj (UNISDR, 2011). Tudi v Evropi se poplave uvrščajo na prvo mesto po pogostosti pojavljanja in predstavljajo kar 75 % vseh odškodnin, ki jih zavarovalnice namenijo poplačilu škode, nastale ob naravnih nesrečah. Posredni ekonomski in družbeni vplivi poplav so še veliko obsežnejši in jih je praktično nemogoče natančno določiti. To kaže na nujnost izvajanja protipoplavnih ukrepov, ki močno zmanjšujejo posledice poplav in z njimi povezano tveganje (Bach in sod., 2004). Nič drugače ni niti v Sloveniji, kjer so poplave v zadnjih letih prav tako

vse pogostejše (Sodnik in Mikoš, 2011). Od leta 2007 smo imeli več obsežnih katastrofalnih poplav. Septembra 2007 so poplave močno prizadele območja Železnikov, Kroke, Bohinja ter območja občin Cerkno in Idrija (Sodnik, 2007; Rusjan in sod., 2009). V juniju leta 2009 so poplave prizadele širše celjsko območje in območje zgornje Save, decembra 2009 območje Bohinja in 11 drugih gorenjskih občin in septembra 2010 kar tretjino ozemlja Slovenije, ko je bilo najhuje v osrednji Sloveniji vključno z ljubljanskim območjem, na območju Celja z Laškim in v Zasavju. Zgornje Posočje, Koroško, območja ob zgornjem toku Savinje, območja ob Dravinji ter območja ob Savi, predvsem v Zasavju in Posavju, pa so poplave ponovno prizadele tudi ob zadnjih poplavah v začetku novembra 2012. Ob takšnih dogodkih se vedno znova pojavi vprašanje, ki ga je Mikoš (2009) izpostavil že v Gostujočem peresu časopisa Delo, to je, ali je pomembneje krepiti sistem zaščite in reševanja ali pa je bolje več vlagati v preventivo, kakovostno poplavalno sanacijo in vzdrževanje vodotokov. Ob omejenih finančnih sredstvih je idealno razmerje med vlaganji v sistem zaščite in reševanja na eni strani in preventivne ukrepe, kakovostno poplavalno sanacijo in vzdrževanje vodotokov na drugi težko določiti. Pozitivni učinki pravilno izvedenih preventivnih ukrepov pred poplavami se zagotovo pokažejo v nižjih stroških zaščite in reševanja. Kakorkoli, zaščitni ukrepi nas pred poplavami ne morejo nikoli povsem zaščititi, zato potrebujemo poleg preventivnih ukrepov za zmanjševanje njihovih posledic tudi dober sistem zaščite in reševanja.

V doktorski disertaciji smo razvili izviren postopek prepoznavanja poplavljenih površin s satelitskih posnetkov s tehnikami strojnega učenja. Cilj je razviti postopek za določitev poplavljenih površin, ki bo omogočal hitrejšo oceno nastale škode, nudil pomoč pri organizaciji reševanja na prizadetih območjih in bil dovolj natančen za poplavalne analize, katerih cilj je določitev vzrokov za nastanek poplav, količine poplavnih voda, poplavno najbolj ogroženih območij ipd. Bach in sod. (2005) so predstavili aplikacijo, ki zagotavlja produkte, kot so karte poplavljenosti in karte vpliva poplav na območjih z različnimi rabami. Vendar se njihova aplikacija omejuje na uporabo satelitskih posnetkov in ročno interpretacijo. Postopek, predstavljen v doktorski disertaciji, kombinira uporabo satelitskih posnetkov z drugimi vrstami podatkov, ki so vsi skupaj analizirani s tehnikami strojnega učenja, s katerimi iz učne množice podatkov pridobimo najpomembnejše informacije, ki vplivajo na zaznavo poplavljenih površin. Postopek je izvedljiv na različnih območjih z možnostjo pojava poplav, kar omogoča izgradnjo modelov za klasifikacijo poplavljenih površin na celotnem območju Slovenije. Več o ozadju problema, metodah dela in ciljih doktorske disertacije sledi v nadaljevanju.

1.1 Ozadje problema

V ozadju doktorske disertacije se skriva precej širši problem kot zgolj klasifikacija poplavljenih površin. Opazovanje poplav zahteva veliko dela, ki obsega pripravo podatkov, pravilno izbiro učnega vzorca in uporabo ustreznega algoritma strojnega učenja. Sledi klasifikacija poplavljenih površin in ocena uspešnosti.

Pri pripravi podatkov je potrebno zagotoviti dober nabor različnih vrst podatkov, ki lahko prispevajo k pravilnejši določitvi poplavljenih površin. Med uporabljenimi podatki so bili satelitski posnetki različnih virov, digitalni model višin (DMV) in mreža vodotokov. Lastnosti, ki jih lahko pridobimo iz teh podatkovnih slojev, imenujemo atributi, s katerimi opisujemo učni vzorec. Učni vzorec sestoji iz primerov, ki čim bolj nazorno opisujejo posamezni klasifikacijski razred, razred poplavljenih in nepoplavljenih površin. Kljub temu da želimo v učno množico podatkov vključiti primere s širokim naborom različnih atributov, moramo biti pozorni na vpliv posameznega atributa in po potrebi preizkusiti učenje z različnimi kombinacijami podatkov. Uporaba odvečnih atributov lahko povzroči napako v modelu in poslabša rezultat klasifikacije. V raziskavi so zato preizkušeni različni vzorci, ki so se razlikovali tako po gostoti učnega vzorca kot po obliki učnih primerov, ki je lahko točka ali območje/segment. Izbrani vzorci so bili nato uporabljeni za izvedbo strojnega učenja, katerega rezultat so bili klasifikacijski modeli za določitev poplavljenih površin. Uspešna izvedba strojnega učenja še ne zagotavlja enake uspešnosti tudi pri klasifikaciji. Klasifikacijo zato opravimo z različnimi modeli, ne samo z modelom, pri katerem je bila uspešnost strojnega učenja največja, in ocenimo uspešnost posameznih klasifikacij z neodvisno množico testnih točk.

1.2 Raziskovalne hipoteze in cilji disertacije

1.2.1 Raziskovalne hipoteze in njihova utemeljitev

Raziskovale hipoteze doktorske disertacije so:

1. Z uporabo strojnega učenja je mogoče izboljšati hitrost in doseči višjo natančnost določitve poplavljenih površin, kot jo zagotavljajo klasični postopki klasifikacije z večspektralnimi satelitskimi posnetki.
2. Relief in pokrovnost tal imata velik vpliv na uspešnost zaznave poplavljenega površja. Vpliv pokrovnosti je mogoče z uporabo strojnega učenja zmanjšati.
3. S časovno vrsto satelitskih posnetkov je mogoče opazovati dinamiko umikanja vode iz poplavljenih površin.

Prva hipoteza predvideva, da uporaba satelitskih posnetkov in tehnik strojnega učenja omogoča hitrejšo in natančnejšo klasifikacijo poplavljenih površin. Dosedanji postopki prepoznavanja rabe tal in pridobivanja drugih informacij iz zemeljskega površja potekajo predvsem na osnovi običajnih postopkov klasifikacije večspektralnih satelitskih posnetkov, kot sta metodi največje verjetnosti (maximum likelihood) ali najbližjega sosedu (K-Nearest). Te metode so običajno uporabljene posredno, preko že izdelanih programskih orodij za nadzorovano klasifikacijo in so zato postopki precej togi tako glede možnosti upoštevanja večjega števila različnih podatkovnih slojev, kot glede natančnejšega prilagajanja učenja posameznemu primeru (npr. oblike učnega vzorca). V naši raziskavi

bomo namesto togih postopkov klasifikacije, pri katerih faza učenja poteka po strogo določenem postopku, postopek učenja in postopek klasifikacije razdelil na dva povsem samostojna dela. Postopek učenja bo tako potekal z uporabo algoritmov strojnega učenja v specializiranih programih namenjenih strojnemu učenju, ki omogočajo prilagajanje učenja konkretnemu primeru. Strojno učenje bo potekalo z večjim številom različnih podatkovnih slojev in različnimi učnimi vzorci, kar bo po pričakovanjih povečalo natančnost določitve poplavljenih površin, glede na natančnosti, ki jih dosegamo z običajnimi postopki klasifikacije. Kljub temu, da je hitro kartiranje vroča tema, s katero se ukvarjajo številne raziskovalne skupine po svetu, pa najhitrejši sistemi še vedno ne omogočajo izdelave karte hitreje kot v enem dnevu. V okviru centra odličnosti Vesolje-SI pa razvijamo procesno verigo, ki bo omogočila samodejnost celotnega postopka od zajema posnetka do izdelave karte. Del tega postopka je tudi klasifikacija poplavljenih površin z uporabo strojnega učenja. Čas izdelave karte se bo tako skrajšal na nekaj ur, pri čemer večji del časa še vedno predstavlja obdobje od nastanka nesreče do trenutka pridobitve satelitskega posnetka.

Druga raziskovalna hipoteza se osredotoča na vpliv oblikovanosti reliefa in pokrovnosti tal na zaznavo poplavljenih površin. Relief in pokrovnost tal močno zaznamujeta postopek prepoznavanja poplavljenih površin. Oblikovanost površja vpliva na razlivanje vod po površju, zato je z uporabo digitalnega modela višin (DMV) dobre višinske natančnosti mogoče natančno določiti območja poplavljenosti. Predpogoj dobre višinske natančnosti je zadostna ločljivost DMV, ki pa še ne pomeni tudi zagotovitve zadostne višinske natančnosti, za katero potrebujemo tudi ustrezne izvirne podatke, iz katerih je DMV pridobljen. Dobra natančnost DMV je še posebno pomembna na ravninskih območjih, kjer že majhna sprememba v višini vode povzroči veliko spremembo v obsegu poplavljenega ozemlja. V primeru, ko digitalnega modela višin dobre natančnosti nimamo na razpolago, je potrebno slabše podatke o oblikovanosti površja nadomestiti z uporabo ostalih podatkovnih virov. To so predvsem večspektralni satelitski posnetki z zadostno prostorsko ločljivostjo.

Pokrovnost povzroča težave v primeru visoke vegetacije, ki zakriva poplavljene površine (Van Trung in sod., 2012). Površine prekrte z visoko vegetacijo so na večspektralnih satelitskih posnetkih vidne kot nepoplavljene, zato jih kot takšne obravnavamo tudi v postopku strojnega učenja. Izjema so območja z redkejšo vegetacijo, kjer se skozi njo delno vidijo tudi tla pod vegetacijo. DMV določa višino zemeljskega površja in ne višine vegetacije na njem, zato ne omogoča prepoznavanja visoke vegetacije. Predvideva se, da je z upoštevanjem preostalih podatkovnih virov (hidrologija, večspektralni satelitski posnetki) vključenih v postopek strojnega učenja in z uporabo učnega vzorca segmentov mogoče ločiti območja z visoko vegetacijo od poplavljenih površin in tako izboljšati natančnost določitve poplavljenih površin.

Tretja hipoteza se nanaša na uporabo več zaporednih satelitskih posnetkov nastalih znotraj kratkega časovnega intervala. Predvideva se, da je v primeru časovne vrste posnetkov, ki si sledijo v ustreznem

časovnem zaporedju, mogoče spremljati dinamiko gibanja vod v času (naraščanje ali zmanjševanje obsega poplav) ter oceniti hitrost teh sprememb. Ocena hitrosti umikanja oz. naraščanja vod na poplavljenih površinah je že del hidravlične analize, za katero potrebujemo poleg časovne vrste satelitskih posnetkov tudi DMV ustrezne natančnosti. Analize s časovno vrsto satelitskih posnetkov so ugodne predvsem za kraški tip poplav, pri katerih je proces polnjenja oz. praznjenja površin z vodo počasnejši. Rezultati analiz gibanja vod na poplavljenih površinah so pomembne za nudenje prve pomoči ob poplavah prizadetemu prebivalstvu in za izvedbo ukrepov za izboljšanje poplavne varnosti v prihodnje.

1.2.2 Cilji

Vedno pogostejše ekstremne količine padavin in z njimi povezane poplave povzročajo veliko gmotno škodo. Ta se kaže na kmetijskih površinah, infrastrukturnih objektih, stanovanjskih in gospodarskih poslopih. V najhujših primerih poplave zahtevajo tudi človeška življenja. Z našo raziskavo želimo:

- razviti hiter postopek prepoznavanja poplavljenih površin ob nastanku poplav, ki bo zagotavljal pridobitev informacij o stanju na poplavljenih površinah že v nekaj urah po nesreči, s čimer bo izdelek koristen tudi za potrebe reševalnih enot, ki med prvimi nudijo pomoč na prizadetih območjih;
- zagotoviti natančnost določitve poplavljenih površin, ki bo olajšala delo različnim uporabnikom, npr.:
 - pomoč zavarovalnicam pri izplačevanju odškodnin zavarovancem, ki so ob poplavah utrpeli škodo (zmožnost prepoznavanja poplavljenih površin na urbanih območjih - med grajenimi objekti),
 - dodeljevanje ustrezne finančne pomoči za škodo na kmetijskih površinah (zmožnost prepoznavanja poplavljenih površin med kmetijskimi zemljišči z visokimi kulturami),
 - pomoč pri organizaciji reševanja na prizadetih območjih (zmožnost prepoznavanja poplavljenih površin na urbanih območjih - med hišami, stavbami);
- izvesti analizo poplavnega dogodka, ki bi pomagala pri prepoznavanju ključnih vzrokov za nastanek poplav in pri določitvi ustreznih poplavnih ukrepov za preprečevanje hujših posledic ob morebitni ponovitvi tovrstnega dogodka.

Ti cilji doktorske disertacije sledijo tudi smernicam Direktive o obvladovanju poplavne ogroženosti (2007/60/EC), ki od držav članic med drugim predvideva oblikovanje kart poplavne ogroženosti in izdelavo načrtov za obvladovanje ogroženosti (de Moel in sod., 2009).

1.3 Zgradba doktorske disertacije

Doktorska disertacija se prične s predstavitvijo podatkov daljinskega zaznavanja, pregledom stanja na področju prepoznavanja poplavljenih površin iz podatkov daljinskega zaznavanja in kratkim uvodom v strojno učenje (drugo poglavje). Sledi tretje poglavje, katerega namen je natančnejša predstavitev tehnik strojnega učenja. Poglavje nas popelje skozi celoten postopek strojnega učenja, tj. od zbiranja podatkov, definiranja vzorcev za učenje in testiranje, izbire ustreznega algoritma, izgradnje modela s strojnim učenjem, do klasifikacije in ocene uspešnosti. Doktorska disertacija se nadaljuje s četrtnim poglavjem, v katerem sta predstavljeni obe študijski območji in značilnosti poplav, ki se pojavljajo na teh površinah. V petem poglavju sta bila razvita postopka za kartiranje hudourniškega in kraškega tipa poplav. Poglavje v prvem delu prikazuje postopek določevanja poplavljenih površin na primeru poplav v Železnikih 2007 in v drugem delu na primeru poplav na Ljubljanskem barju 2010. Oba dela sta sestavljena iz štirih dodatnih podpoglavij, ki prikazujejo vrsto uporabljenih podatkov, metodologijo dela, rezultate in zaključno razpravo. Doktorska disertacija se zaključi z možnostmi za nadaljnje delo in končnimi ugotovitvami v poglavju Zaključek.

2 UPORABA METOD ZA PREPOZNAVANJE POPLAVLJENIH POVRŠIN

Poglavje v prvem delu podaja nekaj osnovnih informacij o daljinskem zaznavanju in podatkih, ki jih tehnike daljinskega zaznavanja nudijo. Zaključí se s kratkim pregledom stanja na področju prepoznavanja poplavljenih površin z uporabo podatkov daljinskega zaznavanja.

2.1 Daljinsko zaznavanje

Daljinsko zaznavanje je znanost pridobivanja informacij o površju Zemlje, ne da bi z njo prišli v neposreden stik. Pri tem zaznavamo in zapisujemo odbito ali sevano elektromagnetno valovanje (EMV), ga obdelujemo, analiziramo in uporabimo v različnih aplikacijah (Oštir, 2006). Izsevano ali odbito energijo od zemeljskega površja zaznavamo s senzorji, ki so pritrjeni na opazovalnih sistemih za daljinsko zaznavanje. Na osnovi zaznanega EMV prepoznavamo objekte in pojave na zemeljskem površju. Zaradi razlik v sodelovanju (interakciji) valovanja različnih valovnih dolžin s površjem (absorpcija, transmisija in odboj) je mogoče razlikovati med različnimi vrstami površja in različnimi objekti.

Objekti in pojavi na zemeljskem površju so lahko najrazličnejših oblik in velikosti. Stopnja podrobnosti, ki jo še lahko zaznamo, je odvisna od prostorske ločljivosti zajetih podatkov, ta pa je odvisna od višine letenja (opazovanje iz vesolja ali zraka) in zmogljivosti snemalnega sistema (senzorja). Količina lastnosti, ki jih lahko določimo objektom in pojavom, je odvisna od spektralnih lastnosti snemalnega sistema in delno tudi od časovne komponente snemanja (sezonski in meteorološki vplivi, kontinuiteta opazovanj). Opazovalne sisteme za daljinsko zaznavanje zemeljskega površja tako lahko delimo na različne načine (Veljanovski in sod., 2012):

- satelitske (Senzorji so na vesoljskih plovilih in satelitih zunaj atmosfere Zemlje.) in zračne (Senzorji so na letalih, helikopterjih, balonih znotraj atmosferskega plašča.),
- optične/pasivne in radarske/aktivne (Optični zajemajo izsevano energijo zemeljskega površja v svetlem delu dneva. Radarski za snemanje uporabljajo lastni vir elektromagnetnega valovanja in so zmožni snemati neodvisno od ure in vremena. Med aktivne sisteme za opazovanje zemeljskega površja se uvršča tudi lidar, katerega način delovanja je podoben delovanju radarja, razlika pa je v uporabljenem delu elektromagnetnega spektra.),
- mono- in večspektralne (Pankromatski in/ali večspektralni zajem. Število spektralnih kanalov je odvisno od lastnosti kamere in namena uporabe podatkov opazovalnega sistema.),
- z enkratnim ali sistematičnim (redno ponavljajočim) načinom opazovanja (sprožena snemanja ali sistematičen velikopovršinski zajem).

Površinske vode imajo razmeroma stalno in stabilno lego v prostoru. Delno se spreminjajo le njihove meje glede na količino vode. Med površinske vode sodijo morja, naravna in umetna jezera, glavni in

stranski vodotoki, potoki, izviri. Te vodne površine v prostoru lahko zato opazujemo s sistematičnimi snemanji. Pri opazovanju poplav pa se običajno srečujemo z enkratnim dogodkom, zato so uporabni predvsem podatki, ki jih pridobimo z enkratnimi snemanji. To so snemanja, ki so sprožena izključno za potrebe opazovanja določenih poplav. Kdaj se bodo poplave pojavile, je praktično nemogoče natančno določiti. Delno je mogoče predvideti le visoke vode, ki se pojavljajo sezonsko, vedno v približno istem obdobju leta. To so npr. narasle vode v presihajočih jezerih ali na močvirnatih območjih. Ostali tipi poplav, npr. hudourniki ali izredne visoke vode, so povsem naključni dogodki in se med sabo lahko zelo razlikujejo. Razlikujejo se lahko npr. po velikosti poplavljenega površja, vrsti rabe na poplavljenem območju, barvi vode, ki preplavlja določeno ozemlje itd. Vse to vpliva tudi na vrsto podatkov daljinskega zaznavanja, ki so v danem trenutku najprimernejši za opazovanje poplav.

2.1.1 Podatki daljinskega zaznavanja

Prednost podatkov daljinskega zaznavanja za opazovanje poplav je, da omogočajo pridobitev informacij obsežnih površin z relativno nizkimi stroški (Schumann in sod, 2009; Mumby in sod., 1999). Dostopnost do prizadetih območij je v primeru naravnih nesreč običajno zelo otežena, z že nekaj posnetki izbranega satelitskega sistema pa lahko pokrijemo celotno opazovano območje. V preteklosti je prevladovalo prepoznavanje poplavljenih površin z radarskih posnetkov. Razlog je bil zlasti v sposobnosti opazovanja skozi oblake, kar v primeru optičnih posnetkov ni mogoče. Zaradi vedno boljše časovne in prostorske ločljivosti pa se povečuje tudi uporaba optičnih večspektralnih posnetkov. Dostopnost optičnih satelitskih posnetkov se je zelo povečala tudi z razvojem malih satelitov, s pomočjo katerih je danes zagotovljena dnevna pokritost katerega koli območja na Zemlji z večspektralnimi posnetki srednje ločljivosti. Za najnatančnejša opazovanja manjših območij potrebujemo visokoločljive satelitske posnetke, aerofotografije in podatke laserskega skeniranja (lidar). Lidarski podatki so še posebej ugodni, ko je prepoznavanje poplavljenih površin oteženo zaradi zaraščenosti rečnih brežin, visokih poljščin, dreves in grmičevja. Težavo različne obarvanosti poplav in vodnih madežev je najlažje reševati z uporabo posnetkov visokih spektralnih ločljivosti. V času oblačnega vremena pa si najpogosteje pomagamo z radarskimi posnetki ali s snemanji z letenjem pod oblaki.

Uspešnost zaznave vodnih površin je zelo odvisna od ustreznosti podatkov. Za izbiro pravih podatkov moramo dobro poznati vrsto senzorjev posameznega opazovalnega sistema in njihove značilnosti. V nadaljevanju zato sledi predstavitev ključnih lastnosti optičnih in radarskih sistemov oz. senzorjev, primernih za prepoznavanje poplavljenih površin.

a) Optični posnetki

Optični senzorji zaznavajo v naravi obstoječo elektromagnetno energijo v vidnem in začetnem infrardečem delu elektromagnetnega spektra, to je valovanje, ki ga oddaja Sonce. Na zemeljskem

površju se valovanje deloma odbije, deloma absorbira in nato ponovno izseva od površja skozi atmosfero. Za odbito Sončevo svetlobo to pomeni, da jo lahko zaznamo samo podnevi, zaradi česar opazovanje ponoči z optičnimi satelitskimi sistemi ni mogoče (Oštir, 2006).

Prednost optičnih posnetkov pred radarskimi je večje število spektralnih kanalov. Različni kanali zaznavajo valovanje različnih valovnih dolžin, to pa prinaša večje število informacij, ki jih lahko uporabimo pri ugotavljanju lastnosti površja, na katerem je prišlo do odboja. Na optičnih posnetkih se vodne površine ločijo od ostalih površin zaradi močnejše absorpcije daljših valovnih dolžin vidne (rdeče) in infrardeče svetlobe ter močnejšega odboja kratkovalovne vidne svetlobe (zelene in modre). Voda je tako običajno izražena v modri ali modrozeleni barvi. Pri opazovanju v rdečem ali infrardečem delu spektra pa je videti temna. V primeru, ko so v vrhnjih plasteh vode raztopljene tudi rudninske snovi, pride do povečanega odboja daljših valovnih dolžin (rdeče svetlobe), zaradi česar je videti svetlejša (Oštir, 2006). Ker se v vidnem delu spektra skozi barvni (intenzitetni) razpon izražajo tudi lastnosti vode, kot so primesi, globina, barvni odsev vodne površine, je klasifikacija praviloma v teh primerih obsežnejša in bolj zahtevna, kot če jo izvajamo v kanalih infrardečega spektra. Izbrane kombinacije spektralnih kanalov različnih valovnih dolžin tvorijo indekse, ki jih prav tako lahko uporabimo pri klasifikaciji. Za prepoznavanje poplavljenih površin je npr. uporaben indeks NDVI (Normalized Difference Vegetation Index) (Frappart in sod., 2006). Vrednosti tega indeksa so na območjih z vegetacijo pozitivne, medtem ko imajo vodna telesa zelo nizke oz. običajno negativne vrednosti.

Na podatkih optičnega daljinskega zaznavanja so vodne površine praviloma dobro prepoznavne. Omejitve lahko predstavljajo prostorska ločljivost, dejanska vidnost in prepoznavnost objektov. Prostorsko ločljivost daljinsko zaznanih podatkov določa razmerje med velikostjo piksla in še vidnimi podrobnostmi na zemeljskem površju. Slabša/nizja ko je ločljivost, več manjših objektov na posnetku bo ostalo neprepoznanih (npr. ozke in majhne njive). Dejanska vidnost se nanaša na objekte, ki nam pri pogledu iz zraka ostanejo skriti ali pa jih ne moremo prepoznati. Tako npr. obrežno rastje zakriva stik vode z bregom. Neprepoznavnost pa predstavlja težavo pri razločevanju objektov, ki imajo zelo podoben odboj EMV – problematično je npr. razločevanje senc in vodnih površin.

Pogosto lahko izbiramo med posnetki, nastalimi s snemanjem iz vesolja, in tistimi, nastalimi z opazovanjem iz zraka z letalom oz. s helikopterjem. Izbiro pogojujejo predvsem želena prostorska ločljivost, značilnosti opazovanega območja in čas zajema. S senzorji, nameščenimi na letalih, med katerimi so vedno pogostejša tudi brezpilotna letala (Laliberte in sod., 2011), lahko pridobivamo prostorsko zelo natančne podatke o zemeljskem površju. Ob sprejemljivih vremenskih razmerah lahko z njimi snemamo kadar koli in na katerem koli delu Zemlje. Njihova omejitev pa je snemanje z nekoliko nižjo spektralno ločljivostjo. Senzorji na letalih beležijo večinoma elektromagnetno valovanje (EMV) v vidnem in izbranem infrardečem delu spektra, medtem ko senzorji na satelitih omogočajo opazovanje v podrobnejši spektralni ločljivosti, pogosto tudi v več infrardečih in termalnih

pasovih EM spektra (večspektralni in hiperspektralni posnetki). To izboljša prepoznavanje in ločevanje različnih geografskih objektov na površju.

b) Radarski posnetki

Radarski posnetki nastanejo z uporabo satelitskih sistemov, ki za opazovanje zemeljskega površja uporabljajo lasten vir elektromagnetnega valovanja v mikrovalovnem delu spektra. Valovanje oddajajo v smeri proti opazovanim predmetom (površju) in beležijo intenziteto odbitega valovanja. Mikrovalovanje lahko prodira skozi oblake, meglice, prah in dež, saj se pri valovanju v daljših valovnih dolžinah izognemo sipanju v atmosferi, ki je izrazito v optičnem delu spektra. Bistveni prednosti radarskih sistemov pred optičnimi sta zato neodvisnost od Sonca kot vira svetlobe, zaradi česar lahko zemeljsko površje opazujemo tudi ponoči, ter zmožnost opazovanja ne glede na vremenske in druge razmere (Oštir, 2006). To je ključno predvsem pri opazovanju in kartiranju dogodkov, ki jih spremlja slabo vreme, med katere spadajo tudi poplave. Omogočajo pa tudi kontinuirana opazovanja stanj in spremljanje spreminjanja vodnih razmer v času.

Za uspešno uporabo radarskih posnetkov moramo poleg osnov delovanja radarskih sistemov poznati tudi način interakcije mikrovalov z zemeljskim površjem. Jakost odbitega radarskega valovanja določa več dejavnikov, med katerimi so najpomembnejši krajevni vpadni kot, razgibanost terena ter prevodnost in dielektričnost zemeljskega površja. Prepoznavanje vode z radarskih posnetkov je izredno ugodno prav zaradi dejstva, da je edini dejavnik, katerega vpliv se na vodnih površinah spreminja, razgibanost površja zaradi valov. Vpliv ostalih dveh dejavnikov (krajevni vpadni kot ter prevodnost in dielektričnost) ostaja bolj ali manj ves čas enak. V primeru, ko je vodna površina povsem gladka, se celotno radarsko valovanje odbije v smeri proč od detektorja. Govorimo o zrcalnem odboju in odbojnost je takrat enaka nič. Vodne površine pa se pokažejo v povsem temni/črni barvi (Henderson in Lewis, 1998; Oštir, 2006). V primerih, ko so na vodni površini zaradi vetra prisotni valovi, govorimo o razpršenem odboju. Šibkejši veter povzroči nastanek drobnih valov, na katerih pride do zmernega odboja, in valovi se pokažejo v nekoliko svetlejših sivinskih odtenkih. Odboj pa je še nekoliko močnejši, kadar se valovi na vodni površini lomijo. Na teh mestih lahko pride do popolnega odboja in valovi se prikažejo v skoraj povsem belih odtenkih.

V idealnih pogojih je zaznavanje vode na radarskih posnetkih razmeroma enostavna naloga. V realnih primerih prepoznavanja poplavljenih površin (Chen in sod., 1999; Puech in sod., 2009; Mason in sod., 2010) pa se pokažejo tudi njihove pomanjkljivosti. Odkrivanje poplavljenosti moti več dejavnikov: predmeti v vodi, »sence« hribov na razgibanem terenu, zrnatost radarskih posnetkov. Zlasti v heterogeno strukturiranih območjih se srečujemo z nezmožnostjo zaznave vode zaradi odboja od različnih objektov na poplavljenem površju. Objekt velikosti že nekaj centimetrov, ki sega nad vodno površino, povzroči, da pride do razpršenega odboja. Del tega valovanja se odbije do sprejemnika. Intenziteta odbojnega signala se poveča in območje ni klasificirano za poplavljeno površino. Takšni

odbojni objekti so lahko npr. hiša, drevo, elektrovod, trs, visoke trave, koruzno polje itd. Razgiban relief pa lahko povzroči tudi, da vrhovi visokih hribov in gora na satelitskih posnetkih zakrivajo pogled na dno doline, kar onemogoča zaznavo poplavljenih površin (Mason in sod., 2012).

c) Lidar

Lidar je aktivni senzor, ki deluje zelo podobno kot radar, vendar v vidnem in infrardečem delu spektra. Razdaljo med senzorjem in opazovanim predmetom določa z merjenjem časa potovanja elektromagnetnega impulza (Oštir, 2006). Z lidarjem trenutno pridobivamo prostorsko najbolj natančne podatke o oblikovanosti površja in njegovi pokrovnosti. Za zelo natančne študije je pomembno znati izbrati ustrezne algoritme obdelave, kot so npr. filtriranje, klasifikacija in rastriranje. Omejitev za uporabnike je lahko velika količina podatkov za obdelavo ter interpretacija zapletenejših upodobitev podatkov.

Odkrivanje vode na izvornih lidarskih podatkih (oblak točk) ali na izdelkih, pridobljenih iz lidarskih podatkov (digitalni model reliefa/površja – DMR/DMP), že v izhodišču ni tako preprosto kot na radarskih in optičnih posnetkih. V oblaku z lidarjem zajetih točk je vidnih veliko podrobnosti zemeljskega površja. Vodna površina zato ni nujno jasno določljiva. Odvisno od lastnosti lidarskega snemanja in senzorja (gostota zajema točk, valovna dolžina) se laserski pulz na vodni gladini lahko odbije (gostota zaznanih laserskih točk na vodni površini je majhna), v plitvi in čisti vodi pa se lahko deloma absorbira in celo odbije od rečnega dna (gostota laserskih točk je povečana). Kljub temu da je lidarsko snemanje največkrat v uporabi za potrebe opazovanja oblikovanosti površja Zemlje na kopnem (Glenn in sod., 2006; Štular in sod., 2012), pa je v uporabi tudi lidar, namenjen opazovanju oblikovanosti površja pod vodnimi površinami plitvih vodnih objektov (npr. reke, jezera, morja) nekje do globine 2 m. Največkrat se pri tem uporablja EMV zelenega dela spektra. Razlikovanje vodne gladine in vodnega dna pa je zaradi majhne razlike med časom potovanja laserskega pulza, odbitega na vodni gladini in na vodnem dnu, ter zaradi v plitvi vodi pogosto kalne vode, lahko zelo oteženo (Pe'eri in Philpot, 2007).

Slovenija načrtuje izvedbo laserskega skeniranja nad svojim celotnim ozemljem (GURS, 2011), kar bo močno olajšalo tudi poplavne analize in določitev poplavno najbolj ogroženih območij. DMV, pridobljen iz lidarskih podatkov, omogoča prepoznavanje pokrovnosti, kot so krošnje dreves, strehe stavb in podobno. DMV, pridobljen iz lidarskih podatkov, smo imeli v času raziskave na voljo le za manjše območje na območju Železnikov v zgornjem delu reke Selške Sore. Za ostala obravnavana območja lidarski podatki niso bili na razpolago.

2.1.2 Pregled stanja na področju prepoznavanja poplavljenih površin z uporabo podatkov daljinskega zaznavanja

V literaturi lahko najdemo veliko uspešnih primerov uporabe satelitskih posnetkov za prepoznavanje poplavljenih površin. Najpogostejši so primeri, ki se nanašajo na prepoznavanje poplav, nastalih ob poplavljalajočih rekah zaradi obilnih padavin, včasih v kombinaciji s taljenjem snega na višje ležečih področjih. Vode se na teh poplavljenih površinah lahko zadržijo še več dni po deževju, zato imamo običajno na voljo dovolj časa za izvedbo snemanj z različnimi satelitskimi sistemi. Na drugi strani so hudourniške poplave za opazovanje s satelitskimi sistemi bistveno zahtevnejše. Pojavijo se veliko bolj nepričakovano, vode pa nato s prizadetega območja tudi hitro odtečejo. Pri teh poplavah se zato pogosto srečamo s pomanjkanjem posnetkov, ki bi bili za opazovanje poplavljenih površin najprimernejši. Za uspešno prepoznavanje poplav si zato pomagamo z različnimi kombinacijami podatkov in različnimi postopki njihove obdelave.

Zaradi dosegljivosti satelitskih posnetkov različnih senzorjev, različnih časov opazovanj, različnih časovnih in prostorskih ločljivosti je združevanje različnih vrst posnetkov vedno pomembnejše orodje v daljinskem zaznavanju (Ayanu in sod., 2012). Na voljo imamo veliko število različnih tehnik združevanja podatkov, ki se med seboj kar precej razlikujejo. Pohl in Van Genderen, ki sta že leta 1998 napravila pregled različnih pristopov, metod in aplikacij za združevanje različnih vrst podatkov daljinskega zaznavanja, izpostavljata naslednja vprašanja, na katera moramo poznati odgovore za izbiro pravilnega pristopa: kakšen je namen naše aplikacije, kakšen tip podatkov potrebujemo, katera je najprimernejša tehnika za združevanje izbranih vrst podatkov, kateri so potrebni koraki predobdelave in katera kombinacija podatkov nam bo najbolj koristila. Najpomembnejše vprašanje, na katero moramo znati najprej odgovoriti, je prvo, saj nam le natančna določitev namena naše aplikacije omogoča izbiro ustreznih podatkov in pravilno izvedbo vseh preostalih korakov.

Liu in sod. (2002) so za opazovanje največjih poplav dvajsetega stoletja, ki so poleti 1998 prizadele severovzhodni del Kitajske, potrebovali aplikacijo, ki je zagotavljala takojšnje in dovolj natančne informacije za določitev poplavljenega ozemlja in nudenje pomoči ob poplavah prizadetemu delu prebivalstva. Z uporabo posnetkov Radarsat so spremljali dinamiko gibanja poplav v različnih časovnih obdobjih. Radarski posnetki so jim omogočili opazovanje celotnega območja zelo obsežnih poplav v vseh vremenskih razmerah in jasno razlikovanje vodnih površin od kopnega. Nato so z uporabo karte pokrovnosti v merilu 1:100.000, ki je bila določena na osnovi posnetkov Landsat TM, za vsako opazovano časovno obdobje določili vrsto poplavljenega ozemlja, kar jim je omogočilo okvirno oceno nastale škode. Primerjava z radarskih posnetkov prepoznanih poplavljenih površin s poplavami, ki so jih določili na osnovi posredovanih podatkov, ki so jih s terena sporočale reševalne ekipe in ob poplavah prizadeti posamezniki (GPS koordinate, fotografije), je potrdila visoko ujemanje. Kljub zelo učinkovitemu spremljanju dinamike poplav, ki ga omogočata uporaba posnetkov z

različnim časom zajema in uporaba različnih senzorjev, pa je njihova določitev poplavljenih površin in ocena škode še vedno potekala z ročno interpretacijo radarskih posnetkov.

Yonghua in sod. (2007) so za določitev vodnih teles (reke, jezera,...), poplavljenih površin in kopnih delov zemeljskega površja na območju vsakoletnih poplav v Maleziji uporabili optične posnetke Landsat ETM iz obdobja z običajnimi vodnimi razmerami in radarske posnetke Radarsat ob poplavah. Posnetki Landsat ETM so omogočili pridobitev informacij o različnih oblikah na zemeljskem površju in velikosti vodnih površin v času običajnih vodnih razmer, posnetki Radarsat pa so prispevali še informacijo o poplavljenih površinah. Za združitev radarskega posnetka in treh IR kanalov posnetkov Landsat ETM 4, 5 in 7 so uporabili tehniko PCA, ki se je izkazala kot zelo učinkovita metoda za kombiniranje tovrstnih podatkov. Končna določitev poplavljenih površin je potekala z nadzorovano klasifikacijo.

Učinkovitost kombiniranja radarskih in optičnih posnetkov se je potrdila tudi v primeru opazovanja poplav, ki so v začetku leta 1995 prizadele osrednji del Nizozemske (Wang in sod., 1996). Med poplavami pridobljeni radarski posnetki ERS 1 so bili uporabljeni v različnih kombinacijah skupaj s posnetki Landsat TM in ostalimi radarskimi posnetki, ki so bili pridobljeni pred poplavami. S kombinacijo kanala TM 1 in kanala TM 4 so določili vodne površine pred poplavami, z dodanim radarskim posnetkom ERS 1, nastalim v času poplav, pa jim je uspelo prepoznati še poplavljenе površine. Kombinacija podatkov TM 1, TM 4 in radarskega posnetka ERS 1 v rdečem, zelenem in modrem kanalu omogoča enostavno določitev vodnih območij, saj se poplavljenе površine prikažejo v svetlo modri in stalne vode v temno modri barvi. S postopki izboljšanja podob, predvsem izboljšanjem kontrasta in z različnimi barvnimi prikazi, pa jim je uspelo razlikovati tudi med travniki, obdelanimi kmetijskimi površinami in urbani območji.

Sandholt in sod. (2003) so napravili preizkus določitve poplavljenih površin s tremi različnimi vrstami posnetkov brez kombiniranja. Za študijsko območje so si izbrali reko Senegal, ki je edini površinski vodni vir severnega dela Senegala. Reka s poplavami pomembno oskrbuje območje z vodnimi zalogami in hranili, ki so pomembna za rodovitnost prsti, na drugi strani pa s poplavami povzroča škodo na infrastrukturi in zahteva tudi človeška življenja. Za opazovanje poplav leta 1999 so uporabili optične posnetke Landsat ETM+ ločljivosti 30 m in 60 m, radarske posnetke ERS 2 ločljivosti 30 m in posnetke AVHRR ločljivosti 1,1 km. Spektralni kanali posnetkov Landsat ETM+ so se izkazali za zelo primerne za prepoznavanje odprtih vod, razgaljenih tal in zelene vegetacije. Razlikovanje med vsemi tremi tipi površja je dobro, do zamenjav prihaja le med vegetacijo in vodnimi površinami v primeru, da slednje vsebujejo tudi veliko plavajočega sedimenta. Razlikovanje vodnih površin z veliko sedimenta in površinami, prekritimi z vegetacijo, pa je uspešno z uporabo IR kanala. Zaradi različnih spektralnih lastnosti vodnih površin, ki so bile posledica vzvalovane vodne gladine, različnih globlin ali pa je bil drugačen odboj posledica poplavljenosti riževega polja, so za njihovo klasifikacijo določili pet različnih klasifikacijskih razredov. Štiri od njih so nato združili v enoten razred poplavljenih

površin, medtem ko je peti razred ostal samostojen in je predstavljal delno potopljena riževa polja pomešana z nepoplavljeno vegetacijo. Nato so poplavljenе površine določili še z uporabo petih radarskih posnetkov ERS 2. Razmik med dvema sosednjima posnetkoma (35 dni) je omogočal spremljanje razvoja poplav v času. Spremljanje poplav pa je bilo mogoče tudi zaradi dlje časa trajajočih poplav in neodvisnosti radarskih sistemov od vremenskih razmer. Pravilnost določitve poplavljenih površin z radarskimi posnetki so potrdili s primerjavo poplavljenih površin prepoznanih s posnetka Landsat ETM+. Nekaj večja odstopanja so se pojavila le na območjih poraslih z vegetacijo, kar je bila posledica spreminjanja višine vegetacije pri spremljanju poplav v času.

Po uporabi posnetkov Landsat in ERS 2 so poplave na istem območju opazovali še z uporabo posnetkov NOAA AVHRR. Njihova glavna pomanjkljivost je slabša prostorska ločljivost ($1,1 \times 1,1$ km v nadirju), zaradi katere veliko pikslov predstavlja površje, ki je le deloma poplavljeno, deloma pa je kopno. Za natančnejšo določitev poplavljenih površin je zato potrebna uporaba informacij s podpikselsko natančnostjo. Časovne in spektralne lastnosti posnetkov AVHRR so za opazovanje poplav bolj ugodne. Pogosti prehodi satelita omogočajo pridobitev posnetkov brez oblačnosti tudi v času pogostih padavin, dva termalna kanala pa nudita zaznavanje temperaturnih razlik, ki so lahko med vodnimi in kopnimi površinami izrazite. Raziskava je pokazala, da posnetki AVHRR ob uporabi metod podpikselskega prepoznavanja zagotavljajo dobro oceno poplavljenosti. Napačno so občasno kot poplavljenе površine prepoznana le območja porasla z vegetacijo, kar je posledica podobnih temperatur na območjih pokritih z vegetacijo in vodnih površinah. Pomoč v teh primerih predstavlja upoštevanje drugega kanala, ki zagotavlja informacijo o zdravju vegetacije oz. stopnji pozelenelosti. V tem kanalu se namreč valovanje na območjih z vegetacijo močneje odbija in na vodnih površinah močneje absorbira.

Frappart in sod. (2006) so ob spremljanju poplav na območjih ob reki Mekong poleg poplavljenega ozemlja ocenjevali tudi količino vode. Za določitev poplavljenih površin so uporabili indeks NDVI, ki so ga izračunali z uporabo vegetacijskega senzorja VGT, nameščenega na satelitu SPOT 4. Kot poplavljena so ocenili tista ozemlja, katerih vrednost NDVI je bila nižja od 0,2. Največ težav za določitev poplavljenih površin so imeli zaradi prisotnosti oblakov. Oblačnost in slabi pogoji v atmosferi znižajo vrednost indeksa NDVI, zaradi česar je količina poplavljenih površin lahko precenjena. Nato so z radarskimi sistemi Topex/Poseidon, ERS 2 in Envisat določili še višino vode na poplavljenih površinah, kar jim je omogočilo spremljanje volumna vode. Njihova raziskava prikazuje nov način za določanje volumnov vode obsežnejših letnih poplav. Takšne meritve pa ne služijo samo spremljanju in napovedovanju poplav, pač pa tudi aplikacijam, kot sta npr. ocenjevanje količine prenesenega sedimenta ali ocenjevanje vodnih zalog v tleh.

Stancalie in sod. (2006) so poleg uporabe optičnih in radarskih satelitskih senzorjev, ki omogočajo opazovanje pojavov na zemeljskem površju (Landsat 7 TM, ASTER Terra, Radarsat), uporabili tudi informacije drugih platform, ki so namenjene opazovanju oblačnosti in drugega vremenskega

dogajanja v nižjih plasteh atmosfere (npr. MODIS Terra, Quikscat DMSP). S temi podatki so dobili informacije o dogajanju v ozračju, ki so nujno potrebne za določitev poplavne ogroženosti. Določevanje poplavne ogroženosti zahteva torej multidisciplinarni pristop, ki vključuje tudi hidrološko/hidravlično modeliranje. Stancalie in sod. za ovrednotenje in upravljanje z vsemi različnimi vrstami podatkov, ki opisujejo dogodek poplav in omogočajo oceno nastale škode, predlagajo uporabo orodij GIS. GIS omogoča pregledno ureditev prostorskih podatkovnih baz (satelitski posnetki, DMV, tematske karte, različni meteorološki in hidrološki parametri). Podatkovne sloje GIS so nato povezali s hidrološko podatkovno bazo, ki jim je omogočila hidravlično modeliranje in izdelavo karte poplavne ogroženosti.

Irimescu in sod. (2010) pri prepoznavanju poplavljenih površin dajejo prednost podatkom, ki omogočajo hitro pridobitev potrebnih informacij. Hitrost kartiranja je pri opazovanju poplav bistvenega pomena, zato veliko pozornost usmerjajo v zadostno časovno ločljivost satelitskih posnetkov. Med satelitskimi posnetki so uporabili posnetke MODIS (Terra in Aqua), ki so kljub slabši prostorski ločljivosti (250–1000 m) zaradi dobre časovne ločljivosti (pokritost celotne Zemlje v 1–2 dneh) koristen podatek za opazovanje zelo obsežnih poplav. Za natančnejšo določitev poplavljenega ozemlja in kontrolo točnosti so bili uporabljeni še posnetki Formosat 2 s prostorsko ločljivostjo 8 m in SPOT 4 s prostorsko ločljivostjo 20 m. Dodatno so bili za tematsko ozadje uporabljeni posnetki Landsat 7 ETM+. Študijsko območje je predstavljala reka Donava, ki je druga najdaljša reka v Evropi. Za sprejemanje pravih končnih odločitev so poleg hitrih potrebni tudi prostorsko dovolj natančne informacije in prijazno uporabniško okolje, v katerem je mogoče kombinirati različne vrste podatkov in imeti celovit pregled nad stanjem v prostoru. V ta namen so podatke daljinskega zaznavanja kombinirali s podatki topografskih kart in karto pokrovnosti (Corine Land Cover 2000). Iz topografskih kart so pridobili sloje administrativnih elementov (državna meja, lokalne skupnosti,...), naravnih elementov (hidrografske mreže: reke, jezovi, kanali, jezera,...) in komunikacijskih omrežij (ceste, železnice). Karta pokrovnosti pa jim je omogočala določitev ekonomske škode. Za določitev poplavljenih površin so izbrali preprost, hiter in hkrati dovolj natančen postopek z uporabo indeksa NDVI. Zaradi zelo nizkih vrednosti indeksa NDVI, ki jih ta dosega na vodnih površinah, je te površine enostavno ločiti od ostalega z vegetacijo pokritega površja. Mejne vrednosti vodnih površin se sicer spreminjajo od posnetka do posnetka, zato jih je potrebno preizkušati za vsak posnetek posebej. Poplavljenе površine so nato pretvorili v vektorski zapis, kar jim je omogočilo lažje prekrivanje z ostalimi sloji in oceno škode. Klasifikacija vodnih površin omogoča tudi razlikovanje med mirnimi, razburkanimi vodnimi površinami ter poplavljenimi zemljišči drugih rab (kmetijska zemljišča, gozd, pašnik). Spremljanje vrste poplavljenе rabe in poplavljenih delov vasi na dnevni ravni jim je omogočilo oceno v poplavah nastale ekonomske škode.

Senzorje na meteoroloških satelitih Terra in Aqua (MODIS) in NOAA (AVHRR) so uporabili tudi Faruolo in sod. (2009). Podobno kot Irimescu in sod. (2010) so tudi oni izpostavili njihovo dobro

časovno ločljivost, ki omogoča izpeljavo informacij v realnem času in sprotno ugotavljanje sprememb. Podatki MODIS pa so zaradi slabše prostorske ločljivosti (250-1000 m) kljub dobri spektralni (36 različnih kanalov) in odlični časovni ločljivosti (za opazovano območje 3-9 ur) uporabni samo za opazovanje velikih območij. Študija avtomatskega določevanja poplavljenih površin in spremljanja razvoja poplav je potekala na primeru poplav v Južni Afriki julija 2007 in Alžiriji oktobra 2008. Dve študijski območji sta omogočili oceno uspešnosti predlaganega pristopa v različnih pogojih opazovanja. Uporabljen je bil pristop RST (Robust Satellite Techniques) (Tramutoli, 2005; Tramutoli, 2007), s pomočjo katerega so na osnovi odstopanj od običajnega stanja, določenega z večletnimi opazovanji, prepoznali poplavljene površine. Za postopek so uporabili spektralni odziv na vodnih površinah v vidnem (VIS) ($0,62 - 0,67 \mu\text{m}$ – kanal 1) in bližnjem infrardečem (NIR) kanalu ($0,84 - 0,88 \mu\text{m}$ – kanal 2). Kanala sta zelo ugodna za prepoznavanje poplavljenih površin, saj se odboj od teh površin razlikuje od odboja na ostalih razgaljenih ali z vegetacijo poraslih površinah. Voda dosega v bližnjem IR delu spektra nižji odboj kot v območju vidne svetlobe. Odbojnosti so upoštevali preko razmerja $R_{\text{NIR}}/R_{\text{VIS}}$ in razlike $R_{\text{NIR}} - R_{\text{VIS}}$, kjer R pomeni odbojnost v izbranem delu elektromagnetnega spektra. Rezultati postopka so pokazali nizko stopnjo napake, kar potrjuje uspešnost metode za avtomatsko prepoznavanje poplavljenih površin v skoraj realnem času.

S pregledom literature smo zasledili veliko različnih kombinacij podatkov daljinskega zaznavanja, ki se uporabljajo za potrebe prepoznavanja poplavljenih površin. Med najpogostejše uporabljenimi kombinacijami satelitskih posnetkov je kombinacija radarskih posnetkov z informacijami o poplavljenih površinah in optičnih posnetkih z informacijami o tipu pokrovnosti pred poplavami. Takšno kombinacijo podatkov, ki omogoča tudi oceno nastale škode, so uporabili Wang in sod. (1996), Liu in sod. (2002), Yonghua in sod. (2007).

Prepoznane poplavljene površine se pogosto uporabijo tudi za potrebe drugih aplikacij, npr. za oceno ob poplavah nastale škode, določevanje stopnje ogroženosti, izboljšanje poplavne varnosti in drugih. Frappart in sod. (2006) so tako karto poplavljenih površin nadgradili še z višinami poplavnih vod, kar jim je omogočilo določitev volumnov vode, ki so za sušna območja afriške celine zelo pomemben podatek. Pogosto se za potrebe opazovanja poplav uporabljajo tudi posnetki slabše prostorske ločljivosti, ki za opazovanje poplav manjših razsežnosti v Sloveniji niso uporabni, v primerih obsežnih poplav pa predvsem zaradi njihove boljše časovne ločljivosti nudijo pomemben vir informacij za spremljanje razvoja poplav v času (Sandholt in sod., 2003; Faruolo in sod., 2009; Irimescu in sod., 2010).

Množičnost uporabe podatkov daljinskega zaznavanja za različne potrebe spremljanja poplav in upravljanja z vodami kažejo na koristnost tovrstnih podatkov in potrebo po nadaljnjem odkrivanju novih možnosti za njihovo uporabo. Prvo in najpomembnejše vprašanje, na katero moramo odgovoriti najprej, če želimo izbrati pravi postopek, je, kakšen je namen naše aplikacije (Pohl in Van Genderen, 1998). Od tega so namreč odvisni tudi izbira podatkov in vsi nadaljnji postopki njihove obdelave.

Pregled stanja je pokazal zlasti potrebo po hitrih in učinkovitih postopkih za prepoznavanje poplavljenih površin. Faruolo in sod. (2009) so tako uporabili postopek RST (Robust Satellite Techniques), ki jim je na osnovi podatkov MODIS nudil avtomatsko prepoznavanje sprememb poplavljenih površin na obsežnih območjih v skoraj realnem času. V številnih primerih pa določitev poplavljenih površin še vedno poteka s pomočjo ročne interpretacije satelitskih posnetkov (Liu in sod., 2002), v kombinacijah s preprostimi postopki različnih barvnih prikazov in izboljšanjem kontrasta (Wang in sod., 1996). Za natančnejšo določitev poplavljenih površin se spektralne lastnosti posameznih kanalov povezujejo še z drugimi podatki, med katerimi je zlasti pogosta uporaba indeksa NDVI (Frappart in sod., 2006; Irimescu in sod., 2010). Rezultati že opravljenih raziskav kažejo na pozitiven učinek združevanja različnih vrst podatkov, ki lahko z dodatnimi informacijami o opazovanem območju zelo pripomorejo k natančnejši določitvi poplavljenih površin. Z večjo količino podatkov pa je povezan tudi daljši čas njihove obdelave, zato potrebujemo učinkovitejše postopke njihove obdelave. V doktorski disertaciji so tako predstavljene tehnike strojnega učenja, ki za potrebe prepoznavanja poplavljenih površin v takšni obliki do sedaj še niso bile uporabljene.

2.2 Strojno učenje

Za prepoznavanje poplavljenih površin s satelitskih posnetkov in iz nekaterih drugih podatkov smo uporabili postopek, sestavljen iz kombinacije strojnega učenja in klasifikacije. Najprej je bilo izvedeno učenje s tehnikami strojnega učenja, ki omogočajo učinkovitejše raziskovanje po podatkih, kot je to v primeru klasičnih postopkov klasifikacije. Klasični postopki klasifikacij s satelitskimi posnetki veliko manj pozornosti namenjajo učenju. Tako npr. pri nenadzorovani klasifikaciji piksle razdelimo v razrede glede na njihovo »naravno« združevanje v spektralnem prostoru. Pri tem predvidevamo, da so vrednosti enake rabe tal (ali podobni predmeti) blizu skupaj v spektralnem prostoru, vrednosti različnih tipov rabe tal pa so daleč narazen. Pri nenadzorovani klasifikaciji tako najprej z upoštevanjem številčnih vrednosti podatkov piksle združujemo v skupine ali gruče, ki jim nato dodelimo ustrezne oznake, ki predstavljajo klasifikacijske razrede, npr. razred poplavljenih in nepoplavljenih površin (Oštir, 2006).

Pri nadzorovani klasifikaciji povezovanje pikslov v skupine ne poteka več samo na osnovi »naravnega« združevanja v spektralnem prostoru, pač pa za ustvarjanje spektralnih vzorcev uporabimo svoje poznavanje zemeljskega površja. Na ta način računalniški program »naučimo«, kaj je poplavljeno in kaj nepoplavljeno ozemlje, nato pa to svoje znanje prenesemo na celoten posnetek. Postopek nadzorovane klasifikacije razdelimo na ustvarjanje in urejanje vzorcev in razvrščanje v razrede (klasifikacija) (Oštir, 2006). Prvi in najpomembnejši korak pri nadzorovani klasifikaciji je izbira vzorcev. Pri tem operater označi območja, kjer ve, da se nahaja določen tip površja. Pri nadzorovani klasifikaciji tako najprej določimo razrede, po katerih želimo izvesti klasifikacijo. Za vsakega izmed razredov moramo izbrati dober vzorec, ki je homogen in dovolj velik. Pri klasifikaciji

oziroma razdelitvi v razrede vsak piksel na podobi primerjamo s poznanimi podpisi v vseh kanalih, pri čemer uporabljamo več primerjalnih oz. klasifikacijskih algoritmov.

Postopek klasifikacije s strojnim učenjem ima vse značilnosti nadzorovane klasifikacije, le da je postopek učenja veliko bolj dovršen in omogoča prilagajanje konkretnemu primeru. Uporaba strojnega učenja v postopku klasifikacije večspektralnih satelitskih posnetkov omogoča pripravo dobrega učnega vzorca, kombiniranje različnih podatkovnih virov in uporabo različnih klasifikacijskih algoritmov za izgradnjo klasifikacijskega modela. Vsi koraki v postopku strojnega učenja so podrobneje predstavljeni v poglavju 3.1. Namen strojnega učenja je učinkovitejše raziskovanje/rudarjenje po podatkih (data mining) in pridobivanje novih, prej neznanih informacij (Maimon in Rokach, 2005). Rezultat strojnega učenja je model, ki omogoča natančnejšo določitev poplavljenih in nepoplavljenih površin, kot s klasičnimi postopki klasifikacije. V uporabi je več različnih tehnik rudarjenja po podatkih, ki se razlikujejo tako po obliki predstavljanja modelov kot po vrsti uporabljenih algoritmov. Tako lahko model predstavimo npr. v obliki odločitvenih dreves (decision tree) (Han in Kim, 2008), odločitvenih pravil (decision rule) (Boutsinas, 2002) ali regresijskih funkcij (regression) (Berk, 2005). Izgradnja modelov lahko poteka ob upoštevanju verjetnostne porazdelitve (Bayesovi klasifikatorji) (Krzysztofowicz, 2001, Krzysztofowicz, 2002). Klasifikacije z modeli, pridobljenimi z uporabo linearnih klasifikatorjev, temeljijo na uporabi linearnih kombinacij različnih lastnosti, ki opisujejo posamezne učne primere (Herbrich, 2001). Lastnosti posameznega primera, predstavljenega v vektorski obliki, lahko uporabimo za učenje z metodo podpornih vektorjev (support vector machines) (Yu in sod., 2006). Vse različne tehnike strojnega učenja pa niso hkrati primerne za reševanje katerega koli problema. Doktorska disertacija se osredotoča na uporabo algoritmov, ki so iz množice učnih podatkov sposobni prepoznati skupne značilnosti posameznega razreda in zgraditi model za klasifikacijo vseh preostalih primerov. Poleg tega so izbrani algoritmi, ki omogočajo izgradnjo modelov v obliki odločitvenih dreves in odločitvenih pravil. Odločitvena drevesa s svojo strukturo zelo nazorno prikazujejo klasifikacijsko pot vsakega posameznega primera in njegovo uvrstitev v izbrani razred. Odločitvena pravila dodelitev vsakega posameznega primera ustreznemu razredu usmerjajo z nizom hierarhičnih pravil.

Glede na obliko učnega vzorca ločimo strojno učenje z uporabo točk in strojno učenje z uporabo segmentov. V primeru učenja z učnim vzorcem točk je klasifikacija potekala na osnovi razvrščanja pikslov v njim pripadajoče razrede. Učenju z učnim vzorcem segmentov pa je sledila objektna klasifikacija. Postopek objektna klasifikacije je sestavljen iz segmentacije in klasifikacije (Veljanovski in sod., 2011c; Blaschke, 2010). Pri segmentaciji podobo najprej razdelimo na segmente, to je skupine pikslov s podobnimi lastnostmi, ki jih določajo atributi, npr. spektralni odboj, velikost, oblika, tekstura in druge. Prednost združevanja pikslov v segmente je možnost določitve večjega števila atributov, predvsem geometrijskih, kar lahko izboljša rezultat klasifikacije (Navulur, 2007). Na drugi strani je piksel zelo majhna enota, ki razen na satelitskih posnetkih nižje ločljivosti ne predstavlja realnega

objekta na zemeljskem površju (Blaschke in Strobl, 2001). Zaradi zaznave prevelikega števila detajlov je pikselska klasifikacija v primerih uporabe podob z visoko prostorsko ločljivostjo običajno neuporabna (Pulvirenti in sod., 2011). Poleg tega naj bi segment že predstavljal zaključen objekt v prostoru, kar omogoča analiziranje medsebojnih razmerij med posameznimi segmenti. Zaradi različnih omejitev (npr. ločljivosti satelitskega posnetka) vsak segment na satelitskem posnetku še ne zagotavlja vedno tudi zaključenega objekta v realnosti. Izmed vseh segmentov izberemo za učenje najprimernejše primere, ki jasno opisujejo lastnosti posameznega razreda in ki bodo zagotavljali razvrstitev v ustrezne razrede tudi preostalih primerov. Obstoječi algoritmi pa ne zagotavljajo segmentacije, ki bi omogočala povsem samodejen zajem vseh iskanih objektov, ampak je potrebna prisotnost operaterja, ki pomaga pri izbiri ključnih objektov (Osman in sod., 2009). Za prepoznavanje so najzahtevnejši majhni objekti s podobnimi spektralnimi in geometrijskimi lastnostmi (Segl in Kaufmann, 2001), ki pa so s pomočjo dodatnih podatkovnih virov, ki jih lahko vključimo v postopek strojnega učenja, lažje določljivi.

3 UPORABA STROJNEGA UČENJA

Zaradi sodobne računalniške in komunikacijske tehnologije količina podatkov, ki je dostopna vse širši množici ljudi, hitro narašča. Ocenjuje se, da se količina podatkov v svetovnih podatkovnih bazah podvoji vsakih 20 mesecev (Witten in Frank, 2005). Tolikšne količine podatkov ne moremo sproti nadzorovati, zato velika večina podatkov ostane v surovi/neobdelani obliki. V želji, da bi čim manj potencialno pomembnih informacij ostalo skritih med podatki, si pomagamo z različno tehnologijo. Sodobna oprema omogoča avtomatsko (oz. pogosteje polavtomatsko) obdelavo podatkov in njihovo shranjevanje v digitalni obliki. Pri tem uporabljamo poleg številnih klasičnih statističnih metod za analiziranje podatkov tudi tehnike strojnega učenja.

Strojno učenje (machine learning) je tehnika avtomatskega učenja. Obsega veliko skupino algoritmov, ki so iz obsežne množice podatkov sposobni prepoznati značilne vzorce oz. njihove skupne značilnosti (Alpaydin, 2010). Vzorce, ki jih pridobimo oz. se jih naučimo iz podatkov za učenje, lahko nato koristno uporabimo za nadaljnje analize in sprejemanje pametnih odločitev. Namen strojnega učenja je torej samodejno pregledovanje podatkov in iskanje njihovih bistvenih značilnosti. Vseh podatkov ne moremo nikoli v celoti pregledati sami, zato si pomagamo z orodji, ki to omogočajo (skoraj) samodejno. Raziskovanje se tako prične na manjšem delu podatkov – vzorcu, katerega namen je določitev splošnih ugotovitev oz. izgradnja modelov, ki veljajo za celotno množico podatkov. Modeli nato s svojim samodejnim delovanjem omogočajo analizo večjih količin podatkov. Takšne analize so nujno potrebne za razumevanje najrazličnejših pojavov, ki nas obkrožajo v vsakodnevnem življenju.

Uporaba strojnega učenja se zaradi sposobnosti obdelave velikih količin podatkov v zelo kratkem času hitro širi na različna področja. Uporaba strojnega učenja je zelo razširjena v ekonomiji (Fabbriatore in sod., 2012; Liao, 2008), medicini (Podgorelec in sod., 2005) ter na številnih drugih področjih, kjer imamo opravka z velikimi količinami podatkov. V doktorski disertaciji smo, glede na nam dostopne vire, prvič uporabili strojno učenje na področju prepoznavanja poplavljenih površin iz večspektralnih satelitskih posnetkov. Dosedanji postopki prepoznavanja poplavljenih površin so potekali na osnovi klasičnih postopkov klasifikacije večspektralnih satelitskih posnetkov, to je nadzorovane in nenadzorovane klasifikacije ali vizualne interpretacije. Uporabljeno je bilo torej učenje, ki je večinoma že vključeno v programska orodja. V naši raziskavi pa smo postopek učenja in postopek klasifikacije razdelil na dva povsem samostojna dela. Postopek učenja je potekal s preizkusom različnih klasifikacijskih algoritmov v namenskih programih za strojno učenje CLUS in WEKA, »klasične« programe za obdelavo satelitskih posnetkov pa smo uporabili zgolj za pripravo podatkov in kartiranje. Za primerjavo natančnosti določitve poplavljenih površin, ki so bile pridobljene s programoma za strojno učenje, smo na območju Ljubljanskega barja poplavljenе površine določili tudi z običajnim postopkom klasifikacije, preko v programu ENVI pripravljenega programskega orodja Feature Extraction.

Veliko človeškega dela v celotnem postopku prepoznavanja poplavljenih površin s tehnikami strojnega učenja zahteva začetek postopka, ki obsega pripravo podatkov in izbiro ustreznega učnega vzorca za izgradnjo modela prepoznavanja poplavljenih površin (Weiss in Provost, 2003). Sledi faza učenja, ki poteka (pol)avtomatsko, vendar se v postopku lahko vseeno srečamo s kar nekaj težavami, ki jih moramo znati pravilno reševati. Mnogi učni vzorci so vsakdanji ali nepomembni. Drugi lahko vsebujejo napake, ki se v vzorcu pojavijo povsem po naključju. Dejstvo je namreč, da podatki iz resničnega sveta niso nikoli idealni. Nekateri podatki bodo vedno nepotrebni, drugi bodo pomanjkljivi. Torej tudi vsa nova odkritja ne bodo nikoli povsem točna. K vsakemu na novo oblikovanemu pravilu se lahko najde izjema, ki mu ne bo ustrezala. Algoritmi morajo biti zato dovolj robustni, da omogočajo delo z neidealnimi podatki in so sposobni izluščiti pravila, ki sicer ne bodo povsem točna, bodo pa vseeno uporabna.

Tehnike strojnega učenja omogočajo tehnično osnovo, s katero je mogoče prepoznavanje vzorcev iz podatkov, ki se lahko uporabijo za napovedovanje, razlaganje in razumevanje dogodkov. Algoritmi napovedovanja so sposobni iz podatkov, ki opisujejo posamezne primere ali dogodke iz preteklosti, predvidevati, kaj se bo zgodilo v podobni situaciji v prihodnosti. Enako pomembni so tudi algoritmi, katerih rezultat učenja je opis trenutnega stanja, ki ga lahko uporabimo za klasifikacijo novih primerov. Takšni strukturalni opisi omogočajo tako napovedovanje kot tudi razlago in razumevanje. Možnost vpogleda v zgradbo modelov je glavna prednost modelov, pridobljenih s tehnikami strojnega učenja, v primerjavi s klasičnim statističnim modeliranjem (Witten in Frank, 2005). Na koncu moramo znati pridobljene rezultate s strojnimi učenjem še ustrezno ovrednotiti in postaviti končne ugotovitve.

3.1 Postopek klasifikacije z uporabo strojnega učenja

Z daljinskim zaznavanjem ustvarimo veliko podatkov, saj s posnetki pokrivamo obsežna območja v različnih časovnih obdobjih. Podatki daljinskega zaznavanja so uporabni pri kartiranju pokrovnosti, nadzoru zaraščanja kmetijskih površin, opazovanju gozda, spremljanju naravnih pojavov itd. Pri tem poskušamo na satelitskih posnetkih razlikovati med različnimi vrstami rabe tal in različnimi objekti na zemeljskem površju, pri čemer najpogosteje uporabljamo postopek nadzorovane klasifikacije. Nadzorovana klasifikacija predstavlja razvrščanje posameznih primerov/pikslov v predhodno določene razrede, ki jih je potrebno opisati z določenim številom učnih primerov. Učna množica je potrebna za nabor lastnosti, ki opisujejo posamezen razred. Preostali primeri se razvrščajo v ustrezne razrede na osnovi podobnih značilnosti. Lastnosti, ki opisujejo posamezen razred in po katerih se ta razred značilno razlikuje od preostalih razredov, določimo z uporabo tehnik strojnega učenja.

Postopek nadzorovane klasifikacije z uporabo strojnega učenja na satelitskih posnetkih vključuje naslednje korake (Shekhar in sod., 2009):

1. Določitev vzorca s slučajno generiranimi točkami/segmenti.
2. Določitev atributov za učenje.
3. Razdelitev vzorca v množico za učenje in testno množico.
4. Izbira primernega klasifikacijskega algoritma in izvedba učenja.
5. Ocena natančnosti modela na testnih podatkih.
6. Klasifikacija z uporabo izbranega klasifikacijskega modela.

V nadaljevanju sledi kratek opis posameznih korakov v postopku klasifikacije.

3.1.1 Določitev vzorca s slučajno generiranimi točkami/segmenti

Na obravnavanem območju izberemo učni vzorec, ki ga bomo uporabili za izgradnjo modela. Vzorec predstavlja manjši del celotne množice podatkov oz. populacije. Strogih pravil glede velikosti učnega vzorca ni, vendar v splošnem velja, da večji ko je vzorec, tem bolj se približamo dejanskemu stanju v populaciji in izboljšamo uspešnost klasifikacije. Majhni vzorci so precej manj zanesljivi in običajno slabše opisujejo populacijo kot veliki. Kljub temu z velikostjo vzorcev ne pretiravamo, saj prevelik vzorec podaljša potreben čas za obdelavo podatkov, poleg tega pa lahko prisotnost odvečnih podatkov povzroči znižanje uspešnosti klasifikacije (Sivapragasam in Liong, 2004). V doktorski disertaciji je učenje potekalo z uporabo učnih primerov, ki pripadajo znanemu razredu, tj. razredu poplavljenih oz. nepoplavljenih površin. Okvirno oceno velikosti takšnega učnega vzorca na satelitskih posnetkih navaja Shekhar in sod. (2009), ki predvideva, da naj bi število učnih primerov za posamezen razred znašalo od 10 do $100 \times$ št. kanalov. Zbiranje učnih primerov z znano vrednostjo pripadajočega razreda pa je zamudno, zato se uveljavljajo tudi tehnike za učenje z neoznačenimi učnimi primeri, tj. z učnimi primeri, za katere ne vemo, kateremu razredu pripadajo, vendar pa v kombinaciji z označenimi učnimi primeri pripomorejo k boljši klasifikaciji (Mitchell, 1999; Goldman in Zhou, 2000; Nigam in sod., 2000).

3.1.2 Določitev atributov za učenje

Izbor kvalitetnih podatkov za učenje je predpogoj za uspešnost aplikacije, v danem primeru za prepoznavanje poplavljenih površin. Najprej premislimo, kateri podatki so pomembni za uspešno rešitev obravnavanega problema. Nato pridobimo vse dosegljive podatke in iz njih izpeljemo potrebne attribute. Odsotnost kakšnega pomembnega podatka pomeni neupoštevanje vseh dejavnikov, ki sooblikujejo opazovani proces. V postopek učenja zato vključimo vse podatke, ki bi lahko imeli določen vpliv na dogodek. Algoritmi strojnega učenja nepomembne attribute izločijo iz nadaljnjega postopka, atributom, vključenim v postopek klasifikacije, pa določijo stopnjo pomembnosti. Vseeno v

postopek strojnega učenja ne vključimo atributov, katerih vpliv lahko a priori (vnaprej) izključimo. Algoritmi strojnega učenja bi namreč lahko v teh atributih po naključju prepoznali določeno povezanost z opazovanim pojavom v prostoru, ki na izbrani učni množici sicer velja, ne velja pa tudi za preostale primere na splošno.

Pri naboru podatkov se srečujemo s številnimi težavami zaradi zbiranja podatkov iz različnih virov. Podatki iz različnih podatkovnih baz se razlikujejo po različnih formatih zapisa, časovnih intervalih obnavljanja podatkov, namenih podatkov in tudi vrstah napak. Kljub temu da je zbiranje podatkov v enoten zapis lahko precej dolgotrajno, pa je nujno potrebno za izdelavo dobrega vzorca, ki bo predstavljal celotno množico podatkov v vsej njeni raznolikosti. Le takšen vzorec lahko uresniči namen oblikovanja dovolj splošnih ugotovitev oz. izgradnjo modela, ki bo veljal za celotno množico podatkov in bo omogočal izpeljavo dovolj natančnih zaključkov tudi iz drugih delov podatkov. Seveda bomo težko našli najboljši vzorec že takoj na začetku, ampak moramo običajno učenje ponoviti z različnimi vzorci in različnimi kombinacijami atributov. Nato primerjamo njihove uspešnosti in izberemo najboljši vzorec.

Ker se doktorska disertacija ukvarja s poplavami, to je s pojavom v prostoru, imamo večinoma opravka s prostorskimi podatki v obliki rastrskih in vektorskih slojev. Podatki oz. atributi učnega vzorca so običajno zbrani v preglednici. V vrsticah so predstavljeni posamezni primeri (objekti) in v stolpcih njihove lastnosti (atributi). Številčni atributi z realnimi vrednostmi so zvezni, atributi z nominalnimi vrednostmi pa diskretni. Primer vzorca z učno množico točk prikazuje preglednica 1.

Preglednica 1: Izsek preglednice z atributi učnega vzorca za prepoznavanje poplavljenih površin ob poplavah v Železnikih 2007. Preglednica prikazuje deset od 255 vrstic/učnih primerov, vključenih v proces učenja.

Table 1: Part of the table with attributes of training samples for determination of flooded areas in Železniki 2007. Table presents 10 out of 255 rows/training samples included in the learning process.

Pr.:	Pop.	B1- 10 m	B2- 10 m	B3- 10 m	B4- 10 m	B1- 2,5 m	B2- 2,5 m	B3- 2,5 m	B4- 2,5 m	Se nadaljuje...
1	0	85	53	70	40	85	53	70	40	
2	0	70	40	44	32	70	40	44	32	
3	1	176	171	94	128	176	171	94	128	
4	1	175	185	84	91	175	185	84	91	
5	0	93	62	69	48	93	62	69	48	
6	1	171	162	77	96	171	162	77	96	
7	0	65	46	71	48	65	46	71	48	
8	1	163	154	81	97	163	154	81	97	
9	0	149	146	107	118	149	146	107	118	
10	0	81	52	47	42	81	52	47	42	

...nadaljevanje.

Pr.:	SPOT-pan	NDVI	NBI	NDBI	Višina	Naklon	Ukriv.	Odd.
1	72	0,148	28,97	-0,284	621,2	37,3	19,2	114
2	60	0,021	35,49	-0,140	707,2	41,1	6,0	128
3	160	-0,293	159,05	0,152	468,1	4, 3	12,4	29
4	183	-0,418	190,06	0,047	468,0	1,8	1,9	23
5	72	0,056	34,89	-0,187	521,0	16,6	4,9	40
6	133	-0,354	163,55	0,114	443,7	0,4	3,4	45
7	70	0,225	39,86	-0,217	545,9	26, 1	22,0	185
8	124	-0,358	158,75	0,111	436,0	0, 3	1,0	40
9	152	-0,072	99,44	-0,055	643,4	8,0	8,6	58
10	62	-0,067	43,43	-0,05	538,9	17,8	31,4	5

Preglednica 1 prikazuje izsek učnega vzorca, ki je bil uporabljen za prepoznavanje poplavljenih površin ob poplavah v Železnikih 2007. Preglednica prikazuje deset od 255 učnih primerov, vključenih v proces učenja, ki so bili uporabljeni v procesu učenja na območju Železnikov. Prvi stolpec predstavlja zaporedno številko učnega primera, drugi stolpec določa iskano spremenljivko poplavljenosti, ostali stolpci vsebujejo vrednosti vseh ostalih atributov. Učni primeri, ki ležijo na nepoplavljenih tleh, imajo vrednost spremenljivke poplavljenosti 0 in učni primeri, ki ležijo na nepoplavljenih tleh, vrednost 1. Ostali atributi so vrednosti učnih primerov, pridobljene iz dveh satelitskih posnetkov SPOT (večspektralni in pankromatski posnetek), digitalnega modela višin in mreže vodotokov.

Uporabili smo večspektralni posnetek s štirimi kanali v ločljivosti 10 m in pankromatski posnetek z enim kanalom v ločljivosti 2,5 m. Prvi posnetek sestavljajo štirje kanali, ki zaznavajo izsevane oz. odbito elektromagnetno valovanje (EMV) od zemeljskega površja. Prva dva kanala zaznavata EMV v območju vidne svetlobe (prvi kanal zeleno svetlobo in drugi kanal rdečo svetlobo) in druga dva kanala EMV v območju infrardeče svetlobe (tretji kanal bližnjo IR in četrti kratkovalovno IR). S kombiniranjem različnih kanalov pridobimo barvne prikaze opazovanega območja. Iz večspektralnega posnetka ločljivost 10 m je bil nato s postopkom izostritve (pan-sharpening) pridobljen tudi posnetek ločljivosti 2,5 m, ki je bil prav tako uporabljen za določitev spektralnih vrednosti učnih primerov. Gre za isti posnetek višje ločljivosti, ki je bil obdelan na Inštitutu za antropološke in prostorske študije ZRC SAZU v okviru aktivacije programa Vesolje in velike nesreče (International Charter Space and Major Disaster). Atributi, ki označujejo količino zaznanega EMV na večspektralnem satelitskem posnetku so označeni z oznakami B1 za vrednosti EMV v zelenem kanalu, B2 za vrednosti EMV v rdečem kanalu, B3 za vrednosti EMV v bližnjem IR kanalu, in B4 za vrednosti EMV v območju kratkovalovnega IR valovanja. Vsak atributi, ki določa vrednost zaznanega EMV v posameznem kanalu večspektralnega posnetka, se ponovi dvakrat, za posnetek ločljivosti 10 m in posnetek ločljivosti 2,5 m. Pankromatski posnetek z enim kanalom zajema območje od zgornjega roba modre svetlobe do rdeče svetlobe (valovne dolžine od 480 nm do 710 nm). Zaznana EMV so na posnetkih

SPOT 5 zapisana kot digitalna vrednost v 8 bitnem zapisu, ki obsega števila od 0 do 255. Z 0 so označena območja brez zaznanega EMV, nižja števila označujejo območja z malo odboja in višja števila območja z več odboja. Število 255 zasedajo tista območja, na katerih je bila zaznana največja količina EMV. Iz posameznih kanalov večspektralnega posnetka so bili izračunani še normiran diferencialni vegetacijski indeks (NDVI), normiran diferencialni indeks grajenega okolja (NDBI) in nov indeks grajenega okolja (NBI). Prvi indeks izkorišča razliko v odboju rdeče in bližnje IR svetlobe. Uporablja se predvsem za opazovanje vegetacije, saj zdravo, živo rastlinje močno odbija svetlobo v bližnjem IR delu EMV, medtem ko rdečo svetlobo močno absorbira. Vrednosti indeksa se gibljejo med -1 in +1, pri čemer indeks dosega višje vrednosti na območjih z bolj intenzivno in zdravo vegetacijo, nižje vrednosti indeksa pa predstavljajo poškodovano vegetacijo in območja brez vegetacije (Oštir, 2006). Druga dva indeksa NDBI in NBI se uporabljata za prepoznavanje pozidanih površin. Indeks NDBI deluje podobno kot vegetacijski NDVI, le da namesto rdeče in bližnje IR svetlobe izkorišča razlike v odboju bližnje IR in kratkovalovne IR svetlobe. Na pozidanih in nerodovitnih zemljiščih se odboj kratkovalovne IR svetlobe glede na odboj bližnje IR močno poveča, medtem ko ostaja odboj na območjih z vegetacijo v obeh kanalih nespremenjen oz. se malce zmanjša, močnejše pa se zniža odboj na vodnih površinah. Vrednosti NDBI višje od 0 označujejo pozidana zemljišča, negativne vrednosti vodna in vrednosti okoli 0 gozdna ter kmetijska zemljišča. Indeks NBI poleg odboja bližnje IR in kratkovalovne IR svetlobe upošteva tudi odboj rdeče svetlobe, kar omogoča ločevanje pozidanih površin tudi od nerodovitnih zemljišč, ki jih z upoštevanjem indeksa NDBI ne moremo popolnoma ločiti.

Atributi pridobljeni iz digitalnega modela višin so bili višina, naklon, usmerjenost in ukrivljenost in so bili izračunani v programu ArcGIS. Atribut višine predstavlja nadmorsko višino v dani točki na DMV in je podana v metrih. Na celotnem obravnavanem območju Železnikov se vrednosti višin gibljejo med 438 m in 737 m. Naklon je izračunan za vsako rastrsko celico in predstavlja največjo spremembo nadmorske višine na razdalji med obravnavano rastrsko celico in tisto sosednjo rastrsko celico v smeri katere je strmina največja. Vrednosti naklona na območju Železnikov so podane v stopinjah in so določene s kotno funkcijo tangens z razmerjem vertikalne in horizontalne spremembe razdalje. Naklon lahko izrazimo tudi v odstotkih, ki ga izračunamo tako, da vrednost količnika med spremembo nadmorske višine in razdaljo med dvema sosednjima rastrskima celicama pomnožimo s 100. Tako npr. sprememba višine za 10 m na razdalji 100 m pomeni 10 % naklon, vrednosti naklona 0 % pa pomenijo povsem ravna območja. S smerjo največjega naklona je določen naslednji atribut usmerjenosti, ki se meri v smeri urinega kazalca in kjer vrednosti 0° in 360° predstavljata usmerjenost proti severu. Ukrivljenost predstavlja stopnjo spremembe naklona na površju in jo izračunamo kot prvi odvod naklona oz. drugi odvod nadmorskih višin.

Zadnji atribut je oddaljenost od vodotokov. Vrednost vsake rastrske celice predstavlja oddaljenost od najbližjega vodotoka, ki so določeni z mrežo vodotokov, izvzetih iz generalizirane kartografske baze v merilu 1 : 25,000 (GKB 25).

Pri določevanju poplav na Ljubljanskem barju so bili podatki učnega vzorca prav tako zbrani v preglednici. Razlike se pojavljajo smo v vrsti posameznih atributov, kar pa je podrobno obrazloženo v poglavju 5.2 Ljubljansko barje - primer kraškega tipa poplav (podpoglavje Podatki).

3.1.3 Razdelitev vzorca v množico za učenje in testno množico

Zbrane podatke razdelimo na dva dela, na podatke za učenje in podatke za testiranje uspešnosti. Podatke lahko ločimo z izbiro dveh območij, od katerih je eno namenjeno izbiri podatkov za učenje in drugo izbiri podatkov za testiranje. Drugi način je izbira obeh vrst podatkov na istem območju. V primeru, ko podatki za učenje in testiranje med seboj niso prostorsko ločeni, je pomembno, da so podatki, ki jih bomo uporabili za testiranje, izbrani neodvisno od podatkov, namenjenih učenju. Okvirno se priporoča dve tretjini podatkov nameniti učenju in preostalo tretjino testiranju uspešnosti. Paziti je potrebno tudi na enakomerno zastopanost primerov, ki pripadajo različnim razredom tako v učni kot testni množici podatkov. Težko je namreč pričakovati, da bo model, ki je bil zgrajen na osnovi učnega vzorca brez prisotnosti primerov določenega klasifikacijskega razreda, sposoben prepoznati primere, ki pripadajo temu razredu (Witten in Frank, 2005).

3.1.4 Izbira primernega klasifikacijskega algoritma in izvedba učenja

Izbira algoritma za izgradnjo modela je odvisna od namena uporabe, za katerega želimo pridobljen model uporabiti. Model vsebuje najpomembnejše elemente dogodka/pojava, ki ga v danem primeru opazujemo. Izogiba se prikazovanju nepomembne vsebine, saj so tako modeli razumljivejši in je iz njih enostavneje razbrati glavne informacije obravnavanega pojava. Za izgradnjo modelov, ki omogočajo razvrščanje posameznih primerov v ustrezne razrede, uporabljamo klasifikacijske algoritme. Med njimi so zelo uporabni algoritmi za izgradnjo odločitvenih dreves (Quinlan, 1986), ki jih delimo na klasifikacijska drevesa za razvrščanje primerov v razrede z diskretnimi vrednostmi in regresijska drevesa za napovedovanje vrednosti zveznih spremenljivk (Quinlan, 1992). Možnost nastavljanja dodatnih parametrov posameznega algoritma omogoča, da delovanje algoritmov še dodatno optimiziramo glede na podatke, ki jih uporabimo v danem primeru. Pri klasifikacijskih algoritmih lahko določimo parametre minimalnega števila primerov v posameznem razredu in način rezanja dreves. Rezanje omogoča skrajšanje dreves, kar zmanjša njihovo razvejanost in povečuje zanesljivost (Chen in Yu, 2007).

3.1.5 Ocena natančnosti modela na testnih podatkih

Modeli so običajno namenjeni reševanju nalog na različnih območjih, na katerih se razmere med seboj lahko zelo razlikujejo, zato potrebujemo trden, kar se da splošen model, ki bo zagotavljal pravilne rezultate tudi v spremenjenih razmerah. Na voljo imamo veliko algoritmov, ki so sposobni skonstruirati model za pravilno klasifikacijo izbrane množice primerov. Preizkusiti pa je potrebno tudi uspešnost njihovega delovanja na drugih primerih z drugačnimi vrednostmi atributov. Katerega od modelov bomo uporabili za končno klasifikacijo, je torej odvisno od ocene natančnosti posameznega modela, ki je bil preizkušen na različnih podatkih. Veliko programov strojnega učenja nudi možnost ocene uspešnosti zgrajenega modela z metodo navzkrižnega preverjanja (*n-fold cross-validation*). Pri tej metodi se podatki učne množice razdelijo na n enako velikih skupin. $(n-1)$ skupin se uporabi za učenje, iz učenja izvzeta skupina pa je namenjena kontroli uspešnosti. Izgradnja modela se nato ponovi n -krat, pri čemer je iz učenja izvzeta vedno druga skupina podatkov, ki je namenjena kontroli uspešnosti. Končen model in ocena natančnosti sta tako rezultat vseh n ponovitev učenja, kar omogoča, da je vsak podatek vsaj enkrat uporabljen tako za učenje kot tudi preverjanje uspešnosti. Vendar pa je ocena, pridobljena iz učne množice podatkov, predvsem ocena uspešnosti učenja in še ne zagotavlja dokončne ustreznosti modela za klasifikacijo. Lahko se namreč zgodi, da se zgrajeni model pretirano prilagaja učnim podatkom in je ocenjena uspešnost previsoka/precenjena. Zato oceno uspešnosti vedno opravimo tudi na od učenja neodvisni testni množici podatkov. Šele na osnovi testov uspešnosti na podatkih, ki za učenje še niso bili uporabljeni, lahko izvemo, kako trden je model v resnici in ali ga je mogoče uporabiti tudi za reševanje ostalih podobnih primerov.

3.1.6 Klasifikacija z uporabo izbranega klasifikacijskega modela

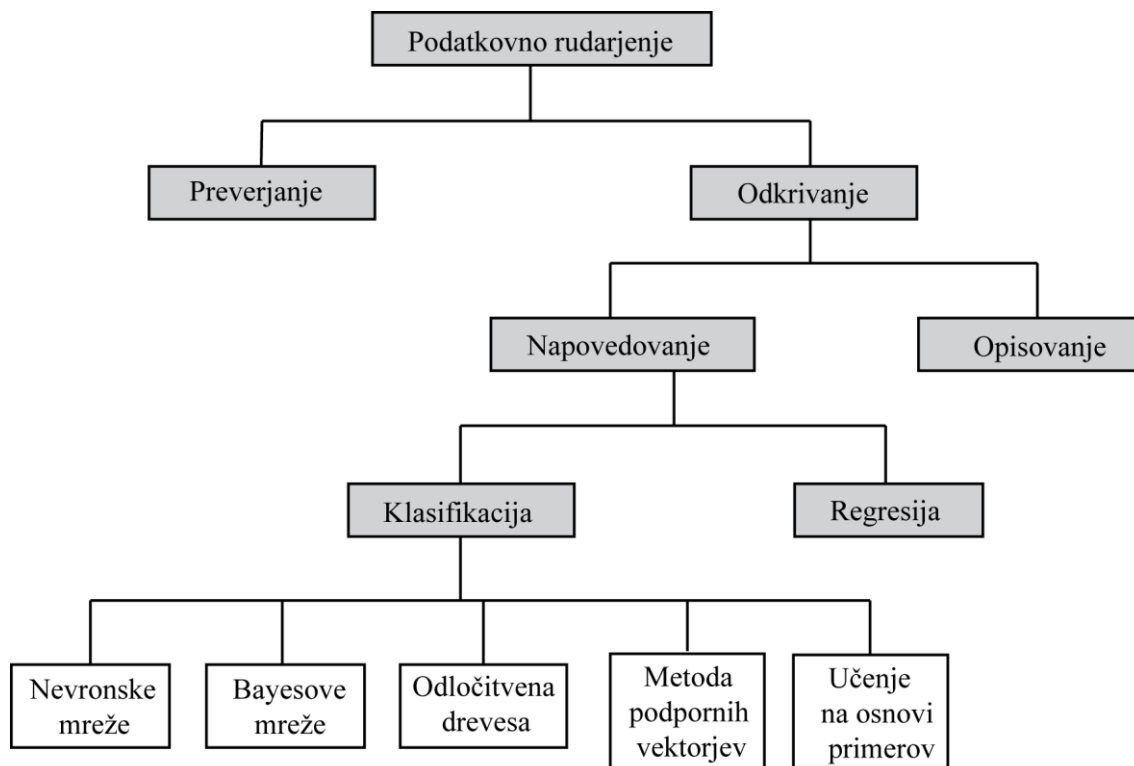
Izmed vseh modelov, ki smo jih pridobili z različnimi algoritmi strojnega učenja (korak 4), izberemo model z najvišjo ocenjeno natančnostjo klasifikacije (korak 5) in ga izvedemo nad celotno množico podatkov. V primeru klasifikacije poplavljenih površin model uporabimo za določitev poplav na celotnem obravnavanem območju. Želja je izgradnja modela, ki bi bil primeren za uporabo tudi ob morebitnih novih poplavah, ne glede na območje njihovega nastanka, vendar pa je ta naloga zelo zahtevna, saj se posamezna območja med seboj zelo razlikujejo po geomorfoloških značilnostih, velike razlike pa se pojavljajo tudi v samem poplavnem dogodku (obsežnost poplav, materialna škoda itd.).

Predvideti je mogoče predvsem strukturo modela in najvplivnejše attribute, ki omogočajo določevanje poplavljenih površin. Vrednosti atributov pa zaradi enkratnosti vsakega poplavnega dogodka in velikih razlik v oblikovanosti posameznega območja vnaprej niso določljive.

3.2 Algoritmi strojnega učenja za klasifikacijo poplavljenih površin

Na voljo je veliko različnih načinov podatkovnega rudarjenja (data mining), ki so namenjeni različnim vrstam uporabe. Osnovna razdelitev podatkovnega rudarjenja razlikuje med metodami za preverjanje že znanih hipotez in metodami za avtomatsko odkrivanje novih pravil in vzorcev iz podatkov (Maimon in Rokach, 2005). V prvem primeru gre v večini primerov za običajne statistične metode, ki so osredotočene na ocenjevanje modelov, npr. test prileganja (fit test) ali analiza odstopanja (analysis of variance). Druga skupina metod je namenjena opisovanju in napovedovanju. Te metode omogočajo odkrivanje novih zvez med podatki ter izgradnjo novih modelov, ki so potem uporabni tudi na drugih še nepoznanih primerih. Metode opisovanja so usmerjene predvsem v interpretacijo in razumevanje podatkov, kar omogoča njihovo medsebojno povezovanje. Primer je npr. metoda gručenja podatkov (clustering) (Estivill-Castro in Yang, 2000). V primerih nenadzorovanega učenja se primeri razvrščajo v skupne razrede brez predhodno določenih ciljnih spremenljivk (Gentleman in Carey, 2008). Ob uporabi nadzorovanih metod pa določamo odnos med vhodnimi atributi (neodvisnimi spremenljivkami) in predhodno določenimi ciljnimi atributi (odvisnimi spremenljivkami). Odnos med njimi prikažemo z modelom, ki je namenjen opisu in razlagi določenega pojava (npr. poplav) in je lahko uporabljen tudi za napovedovanje vrednosti ciljnih spremenljivk ob poznanih vrednostih vhodnih atributov (Kotsiantis, 2007). Razlikujemo med dvema glavnima vrstama nadzorovanih modelov: klasifikacijski in regresijski modeli (slika 1).

Regresijski modeli preslikajo prostor vhodnih podatkov v prostor realnih vrednosti. Regresijo uporabimo, ko želimo pojav opisati z numeričnimi vrednostmi opazovanih lastnosti. Na drugi strani klasifikacija preslika prostor vhodnih podatkov v razrede, ki so določeni z nominalnimi vrednostmi atributa. Zelo primerni obliki za napovedovanje nominalnih vrednosti sta strukturi odločitvenih dreves in odločitvenih pravil, ki sta bili uporabljene v primeru določevanja poplavljenih površin v doktorski disertaciji. Obe obliki sta sicer primerni tudi za napovedovanje numeričnih vrednosti, vendar pa v teh primerih namesto nominalne vrednosti v listih dreves oz. na desni strani pravil stoji povprečna vrednost vseh uporabljenih primerov za učenje. Med zelo uporabnimi algoritmi so še Bayesovi klasifikatorji (Bayesian classifiers) (Heckerman, 1999), ki namesto napovedi razreda podajajo oceno verjetnosti, s katero se posamezen primer uvršča v izbrani razred, vendar doktorska disertacija preizkusa tovrstnih algoritmov, ki za klasifikacijo izkoriščajo verjetnostno porazdelitev, ne vključuje. Za potrebe doktorske disertacije so bili izbrani samo klasifikacijski algoritmi, ki omogočajo neposreden postopek klasifikacije.



Slika 1: Shematski prikaz metod podatkovnega rudarjenja. Zadnja vrstica prikazuje pet primerov postopka klasifikacije: z nevronskimi mrežami, Bayesovimi mrežami, odločitvenimi drevesi, metodo podpornih vektorjev in učenjem na osnovi primerov (Maimon in Rokach, 2005).

Figure 1: Scheme of data mining methods. Last row shows five examples of classifications: neural networks, Bayesian networks, decision trees, support vector machines, instance based learning (Maimon in Rokach, 2005).

V doktorski disertaciji so bili uporabljeni algoritmi za tvorjenje odločitvenih dreves (decision trees) in odločitvenih pravil (decision rules), ki omogočajo razvrščanje primerov, opisanih z numeričnimi atributi, v razred poplavljenih oz. nepoplavljenih površin. Obe metodi sta hitri in učinkoviti. Obliki odločitvenih dreves oz. pravil odnose med posameznimi spremenljivkami prikazujeta zelo nazorno, kar omogoča enostavno razumevanje in izgradnjo modelov za klasifikacijo ostalih primerov. Odločitvena drevesa delujejo po principu razdeli in zmagaj (divide and conquer), kar pomeni, da se celoten problem razdeli na več manjših podproblemov, dokler ti ne postanejo dovolj enostavni za njihovo neposredno reševanje in izgradnjo skupnega modela. Druga tehnika generira pravila, s katerimi ločimo primere, ki pogojem zadoščajo, od primerov, ki jim ne ustrezajo. Vsi uporabljeni algoritmi so predstavljeni v preglednici 2. Iz preglednice je razvidno tudi študijsko območje in vrsta programa, s katerim je bil preizkušen posamezen algoritem strojnega učenja. Zadnja stolpca prikazujeta obliko učnega vzorca, ki ga lahko sestavljajo točke ali segmenti.

Preglednica 2: Algoritmi strojnega učenja, uporabljeni za klasifikacijo poplavljenih površin na območju Železnikov in Ljubljanskega barja. Zadnja dva stolpca označujeta obliko učnega vzorca, ki ga predstavljajo točke oz. segmenti.

Table 2: Machine learning algorithms used for the classification of flooded areas in the area of Železniki and Ljubljana Moor. Two different samples of points and segments were used (signed in the last two columns).

Uporabljeni algoritem	Oblika prikaza	Železniki	Lj. barje	Program	Točke	Segmenti
J48	Odločitveno drevo	X	X	Weka	X	X
JRip	Odločitvena pravila	X	X	Weka	X	
Gozd (Bagging)	Množica odločitvenih dreves	X	X	Weka	X	
Naključni gozd (Random forest)	Množica odločitvenih dreves	X	X	Clus	X	
Podporni vektorji (Support vector machines)	f^*		X	ENVI		X
Najbližji sosed (Nearest neighbor)	f^*		X	ENVI		X

* Algoritma sta vgrajena v postopek klasifikacije z orodjem Feature Extraction v programu ENVI, ki ne omogoča vpogleda v zgradbo modela

Za izgradnjo odločitvenih dreves smo uporabili algoritem J48 in algoritma za kombiniranje klasifikatorjev, to sta gozd (bagging) in naključni gozd (random forest) (Džeroski in Kocev, 2012). Za izgradnjo odločitvenih pravil pa smo uporabili algoritem odločitvenih pravil JRip. Vsako odločitveno drevo gradijo notranja vozlišča in listi. Vozlišča delijo veje dreves na dva dela glede na vrednosti atributov. Učni primeri potujejo skozi vozlišča in se uvrstijo v razred, ki mu pripadajo glede na vse teste v vozliščih od korenin do lista. Listi dreves predstavljajo razrede, ki so v primerih klasifikacije predstavljeni z nominalnimi vrednostmi.

Algoritem J48 uporablja klasifikacijo z enim drevesom. Gozd in naključni gozd pa sta metodi, pri katerih poteka klasifikacija na osnovi večjega števila/množice zgrajenih dreves. Množica dreves (ensemble of classifiers) predstavlja skupino modelov/klasifikatorjev, med katerimi je vsak model zgrajen na osnovi učenja na naključno izbrani podmnožici podatkov. Iz velikega števila modelov je najvplivnejše atribut sicer težje prepoznati, vendar pa predstavlja množica dreves, skupaj s postopkom za njihovo kombiniranje, celoto, ki omogoča pravilnejše predvidevanje. Metodi gozdov in naključnih gozdov sta bili uporabljeni z različnim številom dreves, s čimer smo določili vpliv velikosti množice/števila dreves na natančnost klasifikacije.

V splošnem je težava tako odločitvenih dreves kot odločitvenih pravil, da se zgrajeni modeli lahko pretirano prilagajajo primerom, ki so bili uporabljeni za učenje. To lahko rešimo z rezanjem dreves, s čimer zmanjšamo razvejanost dreves na zadnjih vozliščih in tako pridobimo manj razvejano drevo, ki pravilneje opisuje širšo množico podatkov. Algoritmi, ki ne omogočajo rezanja dreves, so se izkazali kot neprimerni v primeru določevanja poplavljenih površin. Med takšne algoritme se uvršča algoritem

naključnih dreves (random tree), ki je bil preizkušen na študijskem območju Železnikov. Algoritem se je izkazal za neustreznega, saj je rezultat zelo razvejano drevo, ki dobro opisuje učno množico podatkov, na širšem obravnavanem območju pa se njegova natančnost zelo zniža. Vsi preostali algoritmi, ki so bili uporabljeni za določitev poplavljenih površin, omogočajo rezanje. Metodi gozdov in naključnih gozdov poizkušata pretirano prilagajanje modelov učni množici preprečiti tudi z naključno izbiro primerov iz učne množice in naključnim izborom atributov uporabljenih za učenje. To omogoča izgradnjo večjega števila klasifikatorjev (odločitvenih dreves), ki so zgrajeni z uporabo enega učnega algoritma. Napovedi vseh posameznih dreves so nato združene z večinskim glasovanjem, s čimer naj bi se izboljšala točnost in zagotovila večja stabilnost modelov (Ženko, 2003). V našem primeru je bila metoda gozdov uporabljena v kombinaciji z algoritmom J48, ki predstavlja nadgradnjo algoritma C45 (Quinlan, 1993). Izgradnja naključnih gozdov je potekala v programu Clus, kjer je algoritem implementiran v kombinaciji z algoritmom C45.

Med algoritmi za izgradnjo klasifikacijskih pravil smo uporabili algoritem JRip. Izgradnja pravil je bila zaradi binarne klasifikacije poplavljenih in nepoplavljenih površin hitra, saj je pravilna določitev pravil potrebna le za en razred, medtem ko za določitev drugega razreda uporabimo ravno nasprotna pravila.

Za prepoznavanje poplavljenih površin na Ljubljanskem barju sta bili dodatno uporabljeni še metodi podpornih vektorjev (support vector machines) in najbližjega sosedu (nearest neighbor) (Grabowski in sod., 2003). Algoritma sta vključena v postopek klasifikacije s segmenti, ki je potekala z orodjem Feature Extraction v programu ENVI. Postopek omogoča določitev učnega vzorca in izbiro enega od obeh učnih algoritmov. Algoritem podpornih vektorjev omogoča s transformacijami med dvema učnima prostoroma zapletene (nelinearne) primere iz resničnega sveta predstaviti z enostavnimi linearnimi modeli. Algoritem najbližjega sosedu pa razvrsti nov primer v razred glede na najbližje ležeče primere v okolici. V primeru k-najbližjih sosedov se upošteva večje število (k) najbližjih sosednjih primerov, ki so običajno uteženi glede na njihovo oddaljenost.

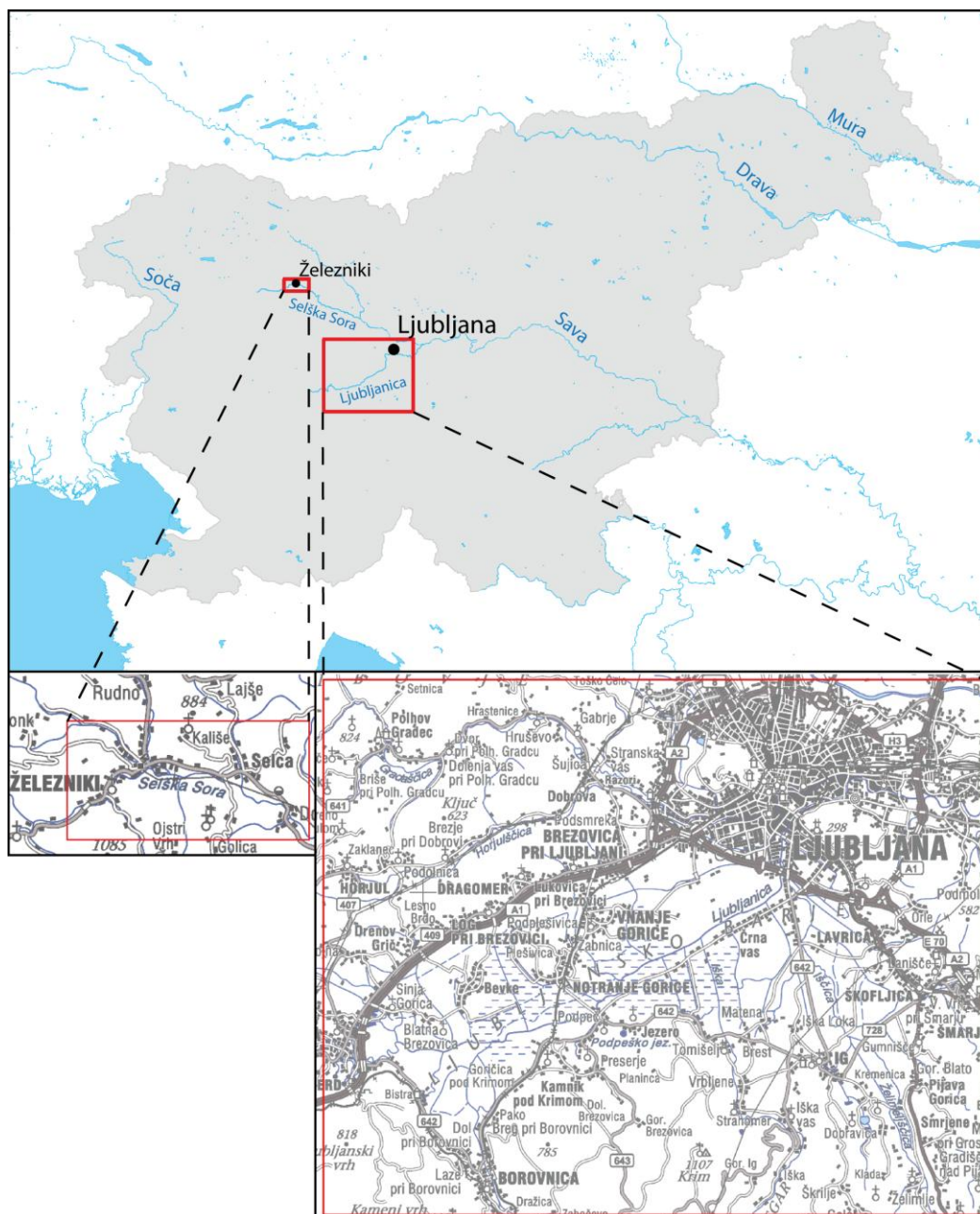
Poleg zgoraj omenjenih algoritmov smo preizkusili še druge, npr. meta algoritem boosting, ki pa se v našem primeru za potrebe klasifikacije niso izkazali za tako uporabne. Meta algoritem boosting z dodeljevanjem uteži posameznim modelom predstavlja bolj zapleten in zato počasnejši postopek klasifikacije (Mease in sod., 2007).

Predstavljeni algoritmi strojnega učenja so bili uporabljeni z uporabo treh različnih programov. Poleg že omenjenega programa za obdelavo satelitskih posnetkov ENVI, s katerim sta potekali klasifikaciji ob uporabi algoritmov podpornih vektorjev in najbližjega sosedu, smo uporabili še dva odprtokodna programa namenjena strojnemu učenju Weka (Hall in sod., 2009) in Clus (Kocev in sod., 2013). Program Clus je izredno dovršen sistem za gradnjo odločitvenih dreves in odločitvenih pravil z uporabo učnih primerov. Uporabljen je bil samo za strojno učenje po metodi naključnih gozdov. V

vseh preostalih primerih je bil uporabljen program Weka, ki vsebuje zbirko algoritmov strojnega učenja za najrazličnejše vrste nalog, s katerimi se lahko srečamo pri raziskovanju po podatkih.

4 ŠTUDIJSKO OBMOČJE

Prepoznavanje poplavljenih površin je potekalo na dveh študijskih območjih: Selški dolini v okolici Železnikov in Ljubljanskega barja. Obravnavani območji sta prikazani na sliki 2. V prvem primeru je bil obravnavan le zgornji del Selške doline, tj. območje naselij Železniki, Selca in Dolenja vas. V drugem primeru je bilo obravnavno celotno območje Ljubljanskega barja vključno z Ljubljano.



Slika 2: Lokaciji študijskih območij: zgornji del doline Selške Sore z Železniki in Ljubljansko barje z Ljubljano. Zgoraj: Prikaz na državnem ozemlju Slovenije. Spodaj: Prikaz na državni topografski karti DTK 50 (GURS, 2005).

Figure 2: The study area location: the upper Selška Sora River valley with Železniki and Ljubljana Moor with Ljubljana. Above: Location in the area of Slovenia. Below: Presentation on digital topographic map DTK 50 (GURS, 2005)

Obe območji se med sabo precej razlikujeta po naravnih značilnostih. Selška dolina je dolga in ozka, po njej teče reka Selška Sora, ki ima hudourniški značaj. V času močnejših padavin količina vode močno naraste in reka dobi veliko erozijsko moč. Obravnavane poplave na območju Selške doline so se zgodile 18. septembra 2007 (Rusjan in sod., 2009; Lamovec in Oštir, 2010). Poleg Selške doline, ki je bila najbolj prizadeta, so poplave prizadele tudi nekatera druga območja na zahodnem in vzhodnem delu Slovenije. Med bolj prizadetimi območji so bila še naselja Cerklje, Bohinjska Bistrica, Kropa in soteska Baške grape na zahodu ter območja ob reki Paki, Bolski in ob spodnjem toku reke Savinje na vzhodu (Pehani in sod., 2008). Dogodek je imel vse značilnosti hudourniških poplav: velika količina padavin ($> 100 \text{ l/m}^2$), kratek čas trajanja padavin (16 ur), velikost poplavnega območja (32–212 km²) in pogostost pojavljanja ekstremnih padavin (izmerjena količina padavin je presegla 100-letno povratno dobo dnevnih padavin, nekatere ocene govore celo o 500-letnih vodah) (Marchi in sod., 2010; Rusjan in sod., 2009). Na prizadetih območjih so poplavljalne številne reke in potoki, poškodovanih je bilo na stotine hiš, odplavljenih veliko mostov in avtomobilov, šest ljudi je umrlo.

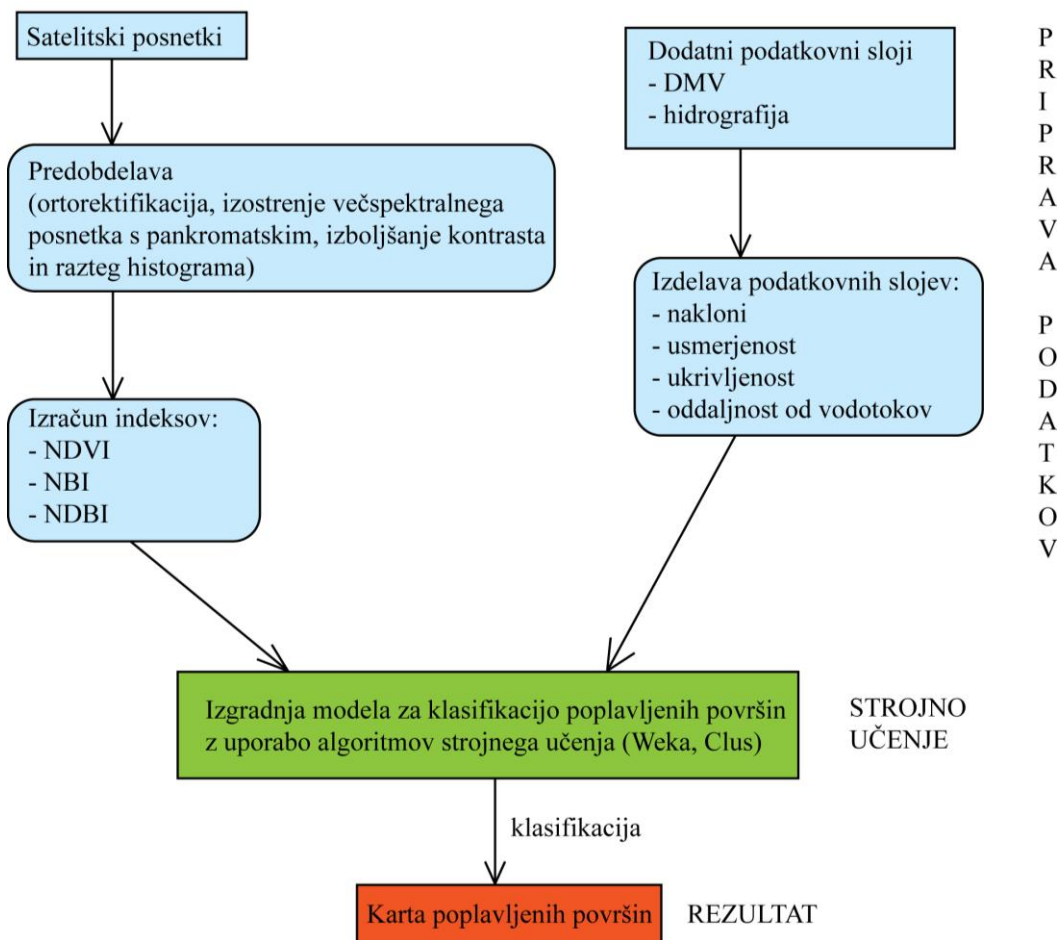
Poplave, ki so se zgodile na Ljubljanskem barju septembra 2010, natanko tri leta po nesreči v Železnikih, so imele povsem drugačne značilnosti. Ljubljansko barje je kraška udornina s prevlado ravninskega sveta. Po njej teče reka Ljubljanica, v katero se stekajo pritoki s kraškimi in površinskimi izviri. Na Ljubljanskem barju so tako poplave kraškega tipa vsakoleten pojav. V nasprotju s hudourniški poplavi je za kraške poplave značilno pogostejše pojavljanje (pojavijo se lahko tudi nekajkrat letno), večji obseg poplavljenega ozemlja, daljši čas trajanja padavin in daljši čas, ki je potreben za umik vode s poplavljenih površin. Pojavljajo se predvsem spomladi ter jeseni, ko se kraška polja ponovno napolnijo z vodami iz ponovno »oživljenih« kraških izvirov. Velika količina padavin v kratkem času in zamašeni požiralniki pa lahko povzročijo, da voda naraste nad običajni nivo. Zaradi izredno majhnega padca in s tem odtoka, voda, ki prestopi robove struge, hitro poplavi zelo obsežne površine. Erozijska moč vode v takšnih poplavih sicer ni tako velika kot v primeru hudourniških poplav, vendar pa je zaradi obsega poplav količina nastale škode lahko vseeno zelo velika. Škoda nastane predvsem na poplavljenih stanovanjskih in gospodarskih poslopih, na poplavljenih kmetijskih površinah, zaliti prometni infrastrukturi ipd. Obilne padavine med 17. in 19. septembrom 2010 so zajele večji del ozemlja Slovenije, poleg Ljubljanskega barja je bilo zaradi poplav prizadetih še veliko drugih območij, zlasti v južnem delu Slovenije. V 48 urah, od petka do nedelje zjutraj, je v povprečju padlo 170–180 mm padavin, kar je največja količina v takšnem časovnem obdobju v zadnjih 60 letih (ARSO, 2010). Dogodek v primerjavi s preteklimi izjemnimi padavinskimi dogodki izstopa predvsem po zelo veliki količini padavin, ki so zajele zelo veliko območje. Visoke vode so najbolj opustošile naselja na območju Ljubljane in Posavja. Povzročile so obsežno škodo v kletih in hišah, poškodovane so bile ceste in železnica, v osrednji in zahodni Sloveniji so se pojavljali številni zemeljski plazovi, umrli so trije ljudje (Veljanovski in sod., 2011b).

5 KARTIRANJE POPLAV

V doktorski disertaciji smo razvili postopek kartiranja poplavljenih površin in ga prikazali na študijskih območjih Železnikov in Ljubljanskega barja. Postopek je sestavni del v okviru Centra odličnosti Vesolje-SI razvijajoče procesne verige, ki bo zagotavljala celotno pot obdelave od surovega optičnega satelitskega posnetka do končnega izdelka, npr. karte poplavljenih površin. Glavni cilj procesne verige je razviti popolnoma avtomatski postopek brez posredovanja operaterja. Procesno verigo sestavljajo modul za geometrijske popravke, modul za radiometrične popravke in modul za izračun NDVI (Oštir in sod., 2012).

Geometrijski popravki so sestavljeni iz treh podmodulov: priprave podatkov, avtomatske določitve veznih točk in ortorektifikacije surovega posnetka. Rezultat prvega modula je ortorektificiran posnetek. Pridobljeni posnetki nastajajo v različnih naravnih pogojih snemanja, zato jih je potrebno radiometrično popraviti, če jih želimo uporabiti za pripravo končnih izdelkov. Popravki vključujejo korekcije zaradi razlik v radiometričnem razponu različnih senzorjev, sprememb v občutljivosti senzorjev, razlik v osvetlitvi in kotu snemanj, razlik v atmosferskih vplivih in razlik v topografiji. Radiometrični popravki tako vključujejo pretvorbo digitalnih vrednosti iz surovih digitalnih vrednosti v vrednosti spektralnega odboja, odpravljanje neidentičnosti odziva detektorjev, atmosferske popravke (zaradi sipanja in absorpcije v atmosferi), topografske popravke (razlike v količini prejete sončne svetlobe zaradi oblikovanosti terena). Rezultat drugega modula sta radiometrično popravljena podoba in maska anomalij (maske oblakov, senc, pikslov brez vrednosti). Tretji modul predstavlja modul za izračun končnih produktov, npr. indeksa NDVI, ki je standardni izdelek v daljinskem zaznavanju za zaznavanje vegetacije in je izračunan iz normirane razlike odboja v rdečem in bližnjem IR kanalu. Distribucija izdelkov končnim uporabnikom bo potekala v obliki spletnega kartiranja. Sistem pa bo poleg spletnih kart podpiral tudi osnovne GIS operacije. Procesna veriga bo nudila uporabniku več različnih izdelkov, ki bodo na voljo za posnetke različnih satelitskih sistemov in za poljubno območje na zemeljskem površju.

Eden od končnih izdelkov bo tudi karta poplavljenih površin, katere postopek izdelave je prikazan v doktorski disertaciji. Postopek določitve poplavljenih površin je prikazan na sliki 3. Postopek se prične s pripravo podatkov, vključenih v postopek strojnega učenja za izgradnjo modela, ki so na sliki 3 prikazani z modro barvo. V raziskavi smo uporabili satelitske posnetke, ki so bili že predobdelani po postopku, predstavljenim s procesno verigo, samostojno pa smo iz spektralnih kanalov satelitskih posnetkov izračunali indekse NDVI, NBI, NDBI, iz digitalnega modela višin (DMV) naklone, usmerjenost in ukrivljenost ter iz hidrografije oddaljenost od vodotokov. Sledila je izvedba strojnega učenja, klasifikacija in izdelava karte poplavljenih površin. Procesna veriga bo zagotavljala predpripravo vseh podatkovnih slojev za celotno območje Slovenije, kar bo omogočalo takojšnjo izvedbo strojnega učenja in klasifikacijo poplav v skoraj realnem času.



Slika 3: Shematski prikaz postopka prepoznavanja poplavljenih površin od priprave podatkov do izdelave karte

Figure 3: The process of flooded areas detection from data preparation to map producing.

DMV in mreža vodotokov nudita dodatne informacije za pomoč pri prepoznavanju poplavljenih površin (Degiorgis in sod., 2012), sama pa ne vsebujeta informacije o dejanskem stanju poplavljenosti. To informacijo vsebujejo samo satelitski posnetki, pridobljeni s snemanjem ob poplavah, zato so slednji ključnega pomena. Satelitski posnetek pa mora nastati v čim krajšem času po nastanku poplav, ko poplave še ohranjajo svoj prvotni obseg in ko voda še ni pričela odtekati. Čas, ki ga imamo na voljo za nastanek posnetka, je zelo odvisen tudi od tipa poplav. Voda v primeru hudourniških poplav odteče zelo hitro. Poplavljenost tla pa ostanejo zelo razmočena tudi po tem, ko je voda že odtekla, kar lahko še vedno zadostuje za zaznavanje poplavljenih površin. Pri določevanju poplavljenega ozemlja so v veliko pomoč tudi sledovi poplav, kot so kamenje in pesek, ki ga je odvrгла deroča voda.

5.1 Železniki - primer hudourniških poplav

5.1.1 Podatki

Za študijsko območje Železnikov smo pridobili veliko količino različnih vrst satelitskih posnetkov, ki so nastali zgolj nekaj dni po poplavah (preglednica 3). Vsi posnetki so bili pridobljeni preko programa

Vesolje in velike nesreče (International Charter Space and Major Disaster). Program je bil aktiviran naslednje jutro po poplavnem dogodku, tj. 19. septembra 2007. Skupaj je bilo posnetih 25 satelitskih posnetkov (Pehani in sod., 2008). Izmed vseh snemanj je bilo v raziskavi doktorske disertacije uporabljeno le snemanje s satelitskim sistemom SPOT 5, ki se je izkazalo kot najbolj uporabno za prepoznavanje poplav. Satelitski sistem SPOT je zagotovil dva posnetka območja v dveh načinih snemanja, pankromatsko in večspektralno, ki sta nastala ob istem času. Pankromatski posnetek ima ločljivost 2,5 m in večspektralni posnetek pa 10 m, ki mu je bila nato s postopkom izostritve (pan-sharpening) ločljivost izboljšana na 2,5 m, s čimer smo pridobili še niz atributov večspektralnega satelitskega posnetka višje ločljivosti. Optični satelitski posnetki SPOT so se kot zelo učinkoviti izkazali že pri analizah plazenja tal (Oštir in sod., 2003; Weirich in Blesius, 2007). Povsem isti posnetki pa so bili prav tako na primeru poplav na območju Železnikov uporabljeni za določitev verjetnostnih razredov pojavljanja plazov pred ujmo in po njej (Jemec in Mikoš, 2008). Na drugi strani radarski posnetki niso mogli biti koristno uporabljeni zaradi hribovitega terena, saj opazovano območje predstavlja del rečne doline, ujete med Škofjeloško hribovje na jugu in predalpsko planoto Jelovico na severu. Gorski vrhovi radarskim senzorjem bočnega pogleda zakrivajo pogled in s tem onemogočajo snemanje dolinskega dna. Tudi uporaba optičnih posnetkov je bila zaradi oblakov in hitrega znižanja vodne gladine ob koncu najsilovitejšega naliva močno otežena. Medtem ko so se satelitski posnetki SPOT kljub temu izkazali za zelo uporabne, posnetkov Formosat in IRS-P6, ki smo jih prav tako pridobili, nismo mogli uporabiti. Uporabo IRS posnetkov je preprečevala njihova nižja ločljivost (23,5 m), ki zadostuje samo za opazovanje večjih območij. Na drugi strani je časovno vrsto štirih posnetkov Formosat preprečevala njihova geometrijska nepravilnost. Nepravilnosti na posnetkih so bile tako velike, da ni bila uspešna niti registracija večspektralnih in pankromatskih parov (Kokalj in sod., 2008).

Uporabljeni posnetki SPOT so bili posneti tri dni po nesreči (21. septembra 2007). Do takrat se je oblačnost že umaknila, prav tako pa se je umaknil tudi glavni poplavni val, zaradi česar je klasifikacija poplavljenih površin potekala na osnovi prepoznavanja poplavnih ostankov, kar je podrobneje opisano v nadaljevanju.

Preglednica 3: Pridobljeni satelitski posnetki za opazovanje poplavljenih površin na območju Železnikov

Table 3: Acquired satellite images for determination of flooded areas in the area of Železniki

Satelit	Ločljivost (m)	Datum	Vir	Vrsta
SPOT 5	ms 10; pan 2,5	21. 9. 2007	CNES	Optični
Formosat	8	21. 9. 2007 22. 9. 2007 23. 9. 2007 24. 9. 2007	NSPO	Optični
IRS P6	23,5	22. 9. 2007	ISRO	Optični
Radarsat	12,5	22. 9. 2007	MDA	Radarski
Envisat	12,5	30. 9. 2007	ESA	Radarski

Atributi, pridobljeni z večspektralnega satelitskega posnetka SPOT 5, so vrednosti posameznih kanalov: zelenega, rdečega, bližnjega IR, srednjega IR kanala ter izračunani indeksi: NDVI (Normalized Difference Vegetation Index) (Burgan in Hartford, 1996), NBI (New Built-up Index) (Chen in sod., 2010) in NDBI (Normalized Difference Built-up Index) (Zha in sod., 2003). Vrednosti teh indeksov so izračunane z naslednjimi enačbami:

$$NDVI = \frac{near_IR - R}{near_IR + R},$$

kjer je NDVI normiran diferencialni vegetacijski indeks (Normalized Difference Vegetation Index),

$$NBI = \frac{R \times SWIR}{near_IR},$$

kjer je NBI novi indeks grajenega okolja (New Built-up Index) in

$$NDBI = \frac{SWIR - near_IR}{SWIR + near_IR},$$

kjer je NDBI normiran diferencialni indeks grajenega okolja (Normalized Difference Built-up Index).

Pri tem je *near_IR* vrednost odboja v bližnjem infrardečem kanalu, *R* vrednost odboja v rdečem kanalu in *SWIR* vrednost odboja kratkovalovnega infrardečega valovanja. Posnetki zaradi hitrejšega postopka določitve poplavljenih površin pred izračunom indeksov niso bili kalibrirani in radiometrično popravljani, zato vrednosti indeksov ne predstavljajo njihovih povsem točnih vrednosti. V okviru razvijajoče procesne verige bodo vsi posnetki radiometrično popravljani, kar bo zagotovilo izračun pravih vrednosti NDVI.

Atributi višin, naklonov in usmerjenost so bili izpeljani iz DMV dveh različnih ločljivosti: DMV 12,5 in DMV 0,5. Model DMV 12,5 je bil izdelan na podlagi metodologije z integracijo različnih podatkovnih virov o reliefu, zajetih od leta 1947 do leta 2005 in pokriva celotno območje Slovenije z okolico (Podobnikar in Mlinar, 2006). Njegova velikost rastrske celice znaša 12,5 m × 12,5 m z višinsko natančnostjo 1,1 m na ravninskih območjih, 2,3 m v gričevnatih in 3,8 m v hribovitih predelih. Na drugi strani DMV 0,5 pokriva le območje najbolj prizadetega zgornjega dela Selške doline (slika 4) in je bil izdelan izključno za potrebe poplavne analize po poplavih v Železnikih. Njegova velikost rastrske celice je 0,5 m × 0,5 m z višinsko natančnostjo 10–30 cm. V primerih uporabe podatkovnih slojev izpeljanih iz DMV 0,5 je bilo obravnavano območje zmanjšano iz celotnega študijskega območja na manjše območje, ki ga pokriva DMV 0,5. Model DMV 0,5 je nastal na osnovi podatkov aero-laserskega skeniranja (ALS), imenovanega tudi lidar (light detection and ranging). Visoka natančnost in gostota podatkov ALS omogočata uporabo različnih tehnik za pridobitev DMV. Prednost laserskega skeniranja pa je tudi možnost pridobitve drugih naravnih in antropogenih podatkov o površju, kot so krošnje dreves, poljščine na kmetijskih površinah, stavbe,

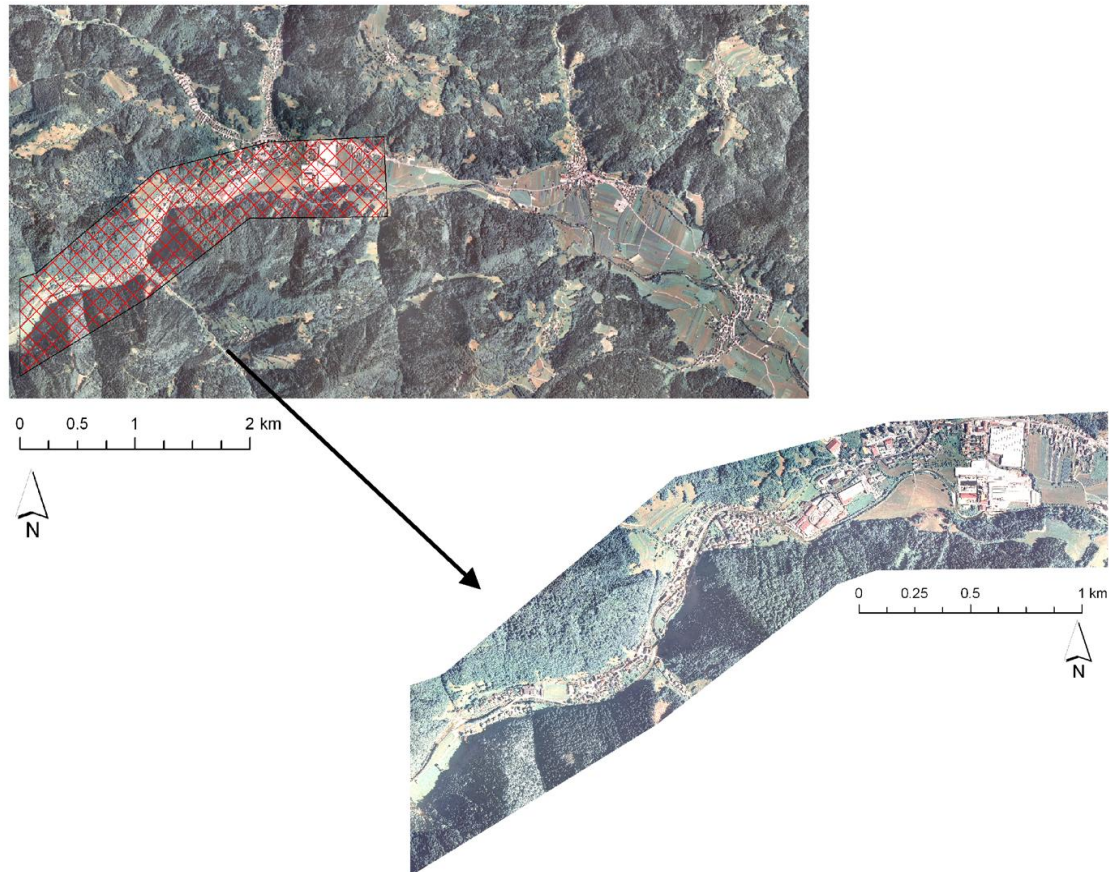
ceste ipd (Podobnikar in sod., 2010). Snemanja s tehnologijo laserskega skeniranja so se v preteklosti v Sloveniji izvajala samo za lokalna območja, na katerih so se pojavile potrebe po tovrstnih podatkih. Trenutno pa je v izvedbi projekt laserskega skeniranja za območje celotne Slovenije, ki bo zagotovil najnatančnejše podatke za celotno državno ozemlje in na osnovi katerih bo izdelan digitalni model reliefa (DMR) z velikostjo celice $1\text{ m} \times 1\text{ m}$ in višinsko natančnostjo 15 cm (GURS, 2011).

Atribut, ki določa oddaljenosti od vodotokov, je bil izračunan na osnovi vektorskega sloja hidrografije generalizirane kartografske baze v merilu 1 : 25.000 (GKB25), ki predstavlja vodotoke, zajete iz topografske karte merila 1 : 25.000.

Preglednica 4: Podatkovni sloji in njihovi atributi, uporabljeni za prepoznavanje poplavljenih površin na območju Železnikov. Pri posameznem kanalu posnetka SPOT 5 so v oklepaju podana spektralna območja zaznanega elektromagnetnega valovanja v nanometrih (nm).

Table 4: Data and their attributes used for determination of flooded areas in the area of Železniki. Wavelength of each spectral band of SPOT 5 sensors is given in brackets in the second column of the table.

Poplave Železniki (2007): 16 atributov	
1.) SPOT 5	
večspektralni (MS): 10 m	B1 (zelena: 500 – 590 nm) B2 (rdeča: 610 – 680 nm) B3 (bližnja IR: 780 – 890 nm) B4 (kratkovalovna IR: 1580 – 1750 nm)
večspektralni (MS), pansharpen/izostren: 2,5 m	B1 (zelena: 500 – 590 nm) B2 (rdeča: 610 – 680 nm) B3 (bližnja IR: 780 – 890 nm) B4 (kratkovalovna IR: 1580 – 1750 nm)
pankromatski (PAN): 2,5 m	SPOT pan (480 – 710 nm)
indeksi: 2,5 m	NDVI NBI NDBI
2.) DMV DMV 12,5 / DMV 0,5	višina naklon usmerjenost
3.) HIDROLOGIJA (omrežje rek)	oddaljenost od vodotokov



Slika 4: Študijsko območje Železniki na orto foto posnetku (GURS, 2006). Zgoraj: Celotno obravnavano območje zgornjega dela Selške doline, ki ga pokriva DMV ločljivosti 12,5 m. Spodaj: Izsek manjšega območja, za katerega je na voljo tudi DMV višje ločljivosti 0,5 m.

Figure 4: The study area of Železniki on orto foto (GURS, 2006). Above: The whole treated area of the upper Selška Sora River valley covered by digital elevation model (DEM) with resolution 12.5 m. Below: Smaller part of the Selška Sora River valley, where DEM 0.5 exists.

Vsi uporabljeni atributi nastopajo v obliki rastrov, ki predstavljajo obliko zapisa podatkov, pri katerih so vrednosti atributov podane v obliki pravih kvadratnih mrež. Točnost rastrske predstavitve podatkov je odvisna od velikosti mrežnih celic, ki jo sestavljajo. Klasifikacije poplavljenih površin so potekale s primerjavo vrednosti atributov posamezne rastrske celice z mejnimi vrednostmi atributov (pragovi), ki so bili določeni v postopku strojnega učenja.

5.1.2 Metode dela

Z namenom čim pravilnejše klasifikacije smo preverili vpliv prostorske ločljivosti digitalnega modela višin, učnega vzorca in vrste algoritma strojnega učenja na natančnost zaznave poplavljenih površin. Učenje je vedno potekalo pod enakimi pogoji, razlika je bila samo v preiskovani lastnosti, kar je omogočilo določitev vpliva te lastnosti na natančnost klasifikacije. Tako se je strojno učenje v primeru preverjanja vpliva prostorske ločljivosti digitalnega modela višin razlikovalo samo v ločljivosti podatkovnih slojev, ki so bili enkrat pridobljeni iz DMV 0,5 in drugič iz DMV 12,5. V primeru

preverjanja vpliva učnega vzorca se je razlika pojavila samo v gostoti učnega vzorca oz. v različni obliki učnih primerov, ki so bili točke oz. segmenti. Pri preizkusu različnih algoritmov strojnega učenja pa se je razlikovala vrsta uporabljenega algoritma. Učni vzorec je bil vedno sestavljen iz 145 učnih točk v primeru učenja na manjšem območju oz. 255 učnih točk v primeru učenja na celotnem obravnavanem območju. Učni vzorec segmentov je bil preizkušen samo na manjšem območju in je bil sestavljen iz 143 segmentov.

S strojnim učenjem pridobljene modele klasifikacije smo zapisali v programski jezik IDL in uporabili za klasifikacijo poplavljenih površin na območju celotnega satelitskega posnetka. Pri klasifikaciji smo obravnavali vsak posamezni piksel, ki je moral vsebovati vrednosti vseh atributov, vključenih v model. Za klasifikacijo piksla v ustrezen razred je bilo tako potrebno vse podatkovne sloje prevzorčiti na enako število pikslov oz. enako ločljivost, ki je bila določena s podatkovnim slojem najvišje ločljivosti. V primeru klasifikacije z uporabo digitalnega modela višin DMV 0,5, je ta hkrati predstavljal tudi sloj z najvišjo ločljivostjo in so bili zato vsi podatkovni sloji prevzorčeni na ločljivost 0,5 m. V primeru uporabe DMV 12,5 so bili vsi sloji prevzorčeni na ločljivost izostrenega večspektralnega satelitskega posnetka SPOT, ki znaša 2,5 m. Vrednosti atributov posameznega piksla so bile nato primerjane z mejnimi vrednostmi oz. pragovi v modelu, ki predstavljajo mejo med dvema razredoma, tj. razredoma poplavljenih in nepoplavljenih površin. Pikslom, ki se uvrščajo v razred poplavljenih površin, se pripiše vrednost 1 in pikslom, ki so uvrščajo v razred nepoplavljenih površin, vrednost 0.

Zaradi preizkusa velikega števila različnih vplivov na uspešnost učenja, je bilo izvedenih veliko različnih kombinacij strojnega učenja, ki so prikazane v preglednici 5 (oznake v tabeli omogočajo sledenja tekom celotne doktorske disertacije). Pri vsakem preizkusu, sta bili opravljeni dve ponovitvi strojnega učenja ki sta se razlikovali v ločljivosti DMV, gostoti učnega vzorca ali obliki učnih primerov. Izjemo predstavlja preizkusa klasifikacijskih algoritmov, kjer so bile opravljene štiri ponovitve strojnega učenja s štirimi različnimi algoritmi. Lastnosti, po katerih so se razlikovale posamezne ponovitve strojnega učenja, so prikazane s krepko poševno pisavo (preglednica 5). V primerih, ko posamezna lastnost ni bila predmet preizkusa, smo za učenje uporabili: DMV 0,5, učni vzorec s 145 učnimi točkami in algoritem J48. Izjemo predstavlja le primer preizkusa vpliva gostote učnega vzorca, kjer je bil namesto DMV 0,5 uporabljen DMV 12,5. Pri oceni vpliva gostote učnega vzorca na natančnost klasifikacije sta bila namreč oba modela preizkušena na celotnem obravnavanem območju, ki pa ga v celoti pokriva le DMV 12,5.

Preglednica 5: Strojno učenje na območju Železnikov. Preglednica prikazuje različne kombinacije strojnega učenja, ki so bile preizkušene glede na ločljivosti uporabljenega DMV, gostoto učnega vzorca, obliko učnih primerov in vrsto uporabljenih algoritmov.

Table 5: Machine learning in study area of Železniki. Table presents different combinations of DEM, training samples and machine learning algorithms which were used for machine learning process.

Vrsta preizkusa	Ločljivost DMV	Št. učnih primerov / velikost območja	Oblika učnega vzorca	Algoritem
Preizkus vpliva prostorske ločljivosti DMV	DMV 12,5 in DMV 0,5	145 točk / manjše območje (58 točk / ha)	točke	J48
Preizkus vpliva učnega vzorca - gostota	DMV 12,5	145 točk / manjše območje (58 točk / ha) in 255 točk / celotno območje (11 točk / ha)	točke	J48
Preizkus vpliva učnega vzorca - oblika	DMV 0,5	145 točk (143 segmentov) / manjše območje (58 točk / ha)	točke in segmenti	J48
Preizkus klasifikacijskih algoritmov	DMV 0,5	145 točk / manjše območje (58 točk / ha)	točke	J48, JRip, gozd in naključni gozd

Ocena natančnosti je v vseh primerih vključevala tako oceno natančnosti učenja kot oceno natančnosti klasifikacije, ki sta bili določeni neodvisno. Za oceno natančnosti učenja smo uporabili postopek 10-kratnega navzkrižnega preverjanja, pri katerem je učni vzorec razdeljen na 10 delov. Pri vsaki od desetih ponovitev strojnega učenja se za učenje uporabi drugih devet delov učnega vzorca, preostali del podatkov pa se uporabi za oceno natančnosti. Končna ocena uspešnosti učenja je rezultat vseh desetih ponovitev strojnega učenja.

Visoka natančnost učenja pa še ne zagotavlja tudi visoke natančnosti klasifikacije, ampak je potrebno oceniti tudi natančnost končne klasifikacije. Vpliv ločljivosti DMV, gostote učnega vzorca, oblike učnih primerov in algoritmov strojnega učenja na natančnost klasifikacije smo tako določili na osnovi natančnosti končne klasifikacije, ki jo zagotavlja posamezen model, pridobljen z določenim postopkom strojnega učenja. Natančnost klasifikacije je bila ocenjena na osnovi testnih točk, ki so bile uporabljene tako v primeru učenja s točkami kot učenja s segmenti. Testne točke so bile enakomerno razporejene po obravnavanem območju neodvisno od učne množice. Za oceno natančnosti klasifikacij, ki so obsegale celotno obravnavano območje, smo uporabili 125 testnih točk in za oceno natančnosti klasifikacij na manjšem izseku obravnavanega območja, 100 testnih točk. V obeh primerih so bile točke razporejene po poplavljenih in nepoplavljenih površinah, njihovo dejansko stanje glede poplavljenosti v času poplav pa je bilo določeno na osnovi vizualne interpretacije večspektralnega satelitskega posnetka SPOT, za pomoč pa smo uporabili tudi obstoječe fotografije in video posnetke,

ki so nastali v času poplav. Ocena tako predstavlja skupno natančnost, ki vključuje tako napake, ki so nastale zaradi napačne klasifikacije poplavljenih površin (podcenjena količina poplavljenih površin), kot napake, ki so nastale zaradi napačne klasifikacije nepoplavljenih površin (precenjena količina poplavljenih površin). Za najuspešnejšo klasifikacijo je bila opravljena tudi podrobnejša analiza natančnosti, podana z izdelovalčevo in uporabnikovo natančnostjo.

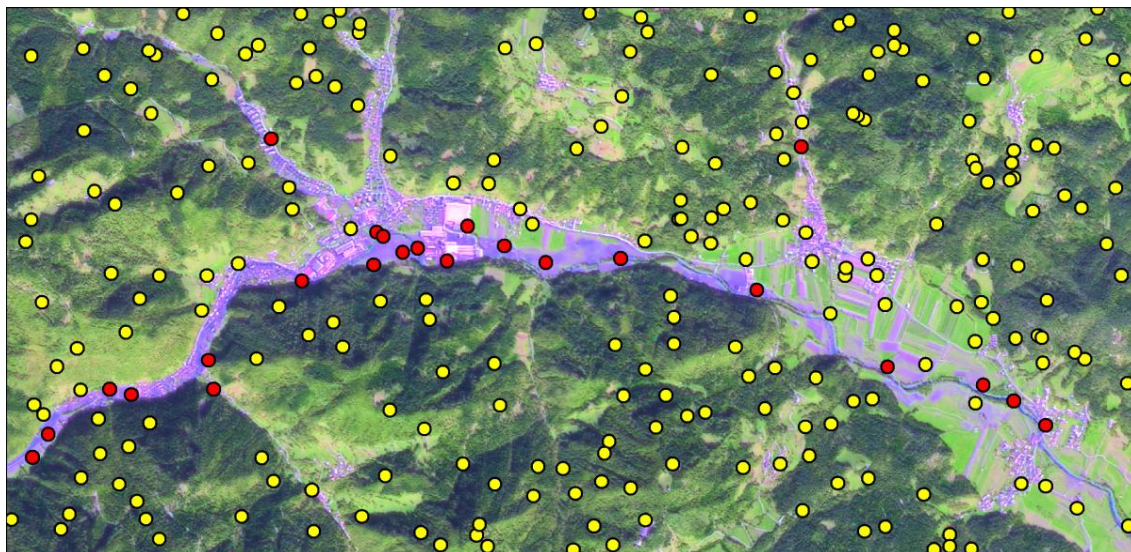
Preizkus vpliva prostorske ločljivosti DMV

V prvem primeru smo preizkusili vpliv ločljivosti DMV na natančnost klasifikacije. Strojno učenje smo izvedli dvakrat, prvič z uporabo digitalnega modela višin slabše ločljivosti (DMV 12,5) in drugič z uporabo digitalnega modela višin višje ločljivosti (DMV 0,5). Vsi ostali pogoji strojnega učenja so bili enaki v obeh primerih. Obakrat je učenje potekalo z algoritmom J48 in naborom 145 učnih točk, ki so bile razporejene po manjšem območju, za katerega smo imeli poleg DMV 12,5 (pokriva celotno Slovenijo) na voljo tudi DMV 0,5 (slika 6). Podatki, ki so bili uporabljeni za klasifikacijo, so bili vedno enaki podatkom, ki smo jih uporabili že v postopku učenja. Tako je v primeru uporabe modela, ki je bil pridobljen z učenjem na DMV 0,5, tudi klasifikacija potekala ob upoštevanju DMV višje ločljivosti in v primeru uporabe modela, pridobljenega z učenjem na DMV 12,5, ob upoštevanju DMV nižje ločljivosti. Obe klasifikaciji sta bili opravljene na manjšem območju, na katerem smo opravili tudi hidravlično analizo, enkrat z upoštevanjem podatkov DMV 0,5 in drugič podatkov DMV 12,5. DMV je bil uporabljen za določitev površine pretočnega prereza, obsega omočenega oboda in naklona dna struge. Glede na lastnosti struge, ki določajo hitrost toka (dno struge, vegetacija ob robu struge in vrsta rabe v okolici) je bil določen še koeficient hrapavosti (Manning). Iz pridobljenih parametrov je sledil izračun hidravličnega radija, povprečne hitrosti in pretoka.

Preizkus vpliva učnega vzorca

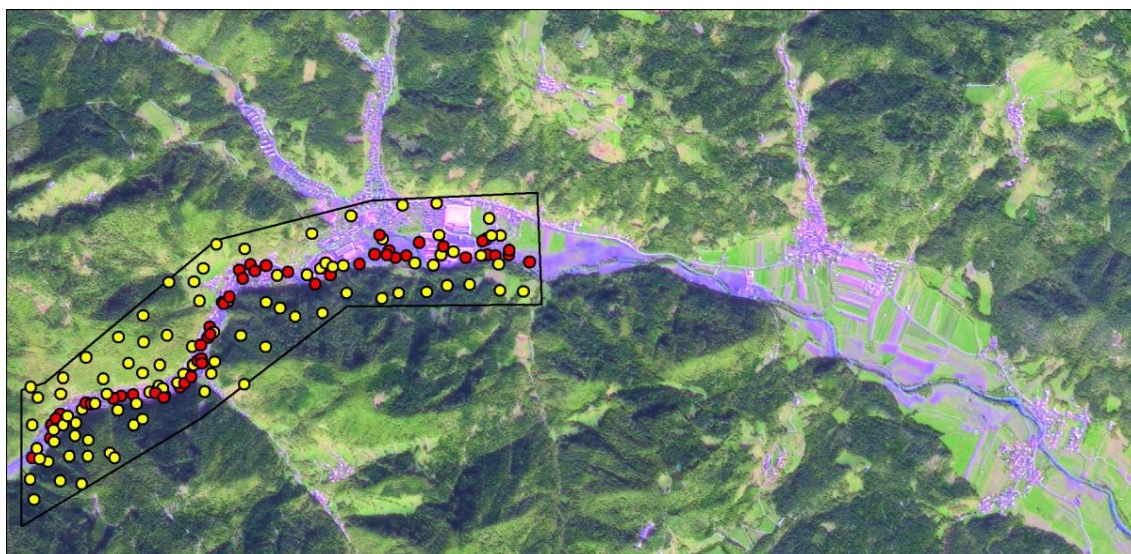
Preizkus vpliva učnega vzorca na natančnost določitve poplavljenih površin smo analizirali z vidika različne gostote učnega vzorca in različne oblike učnih primerov. Gostota učnih primerov je pomembna, ker ta določa stopnjo podrobnosti, s katero učni vzorec opisuje obravnavano območje, oblika učnih primerov pa določa vrsto atributov, s katerimi lahko učni vzorec opisuje obravnavano območje. V primeru preizkusa gostote smo uporabili dva učna vzorca z 255 oziroma 145 točkami. V prvem primeru so bile točke razporejene po celotnem obravnavanem območju (slika 5) z gostoto 11 točk / 100 ha. V drugem primeru pa je bilo 145 točk razporejenih na manjšem izseku (slika 6), s čimer smo dosegli višjo gostoto točk (58 točk / 100 ha). Oba učna vzorca sta bila pridobljena z naključno porazdelitvijo točk. Za določitev vpliva različne gostote učnih točk smo obakrat uporabili enak nabor atributov (obakrat je bil uporabljen DMV 12,5) in isti algoritem učenja J48. Z modelom, ki smo ga pridobili z učenjem na manjšem območju, smo opravili tudi klasifikacijo celotnega obravnavanega območja, s čimer smo preizkusili uspešnost ekstrapolacije modela na širše sosednje območje.

Preizkus oblike učnega vzorca je vključeval učenje z uporabo učnega vzorca sestavljenega iz učnih točk in učnega vzorca sestavljenega iz segmentov. Preizkus je potekal na manjšem območju, ki ga pokriva DMV 0,5 in z uporabo algoritma J48. Učni vzorec točk je vseboval 145 učnih primerov, ki smo ga nato nadomestili s 143 učnimi segmenti (slika 7). Število primerov v učnem vzorcu točk je za dva večje od skupnega števila segmentov, kar je posledica tega, da v dveh primerih na območju istega segmenta ležita dve točki hkrati. Učni vzorec s segmenti smo pridobili s postopkom segmentacije, ki smo jo opravili na izostrenem večspektralnem posnetku SPOT 5 ločljivosti 2,5 m z orodjem Feature Extraction v programu ENVI. V postopku segmentacije se sosednji piksli povezujejo v skupine imenovane segmenti. Povezovanje pikslov v segmente poteka na osnovi iz večspektralnega satelitskega posnetka prepoznanih prostorskih in spektralnih lastnosti ter lastnosti teksture. Velikost segmentov lahko uravnavamo s stopnjo merila (scale level) in stopnjo združevanja segmentov (merge level). V našem primeru smo izvedli segmentacijo s stopnjo merila 10 in brez spajanja segmentov. S tem smo pridobili segmente majhnih velikosti, kar nam je omogočilo, da smo atributom, ki smo jih pridobili med postopkom segmentacije s programom ENVI (42 atributov), pripisali tudi točkovne attribute (12 atributov). Iz razdelitve celotnega območja na segmente smo izbrali segmente, ki so sestavljali učni vzorec za izgradnjo modela za klasifikacijo poplavljenih površin. Tovrsten postopek klasifikacije, pri katerem najprej opravimo segmentacijo in izberemo učne segmente, ki jih nato uporabimo za izgradnjo klasifikacijskega modela, imenujemo objektna klasifikacija (Blaschke in sod., 2008). Skupno število atributov posameznega segmenta je bilo tako 54. Točkovni atributi so bili enaki atributom, predstavljenim v preglednici 1, med katerimi pa smo izpustili vrednosti atributov večspektralnega satelitskega posnetka SPOT ločljivosti 10 m. Atributi segmentov, izračunani med postopkom segmentacije, pa so opisovali njihove geometrijske lastnosti (površino segmenta, dolžino glavne osi, trdnost, izbočenost, konveksnost,...), spektralne lastnosti posameznega kanala večspektralnega posnetka (minimalno, maksimalno in povprečno vrednost posameznega kanala in standardni odklon), teksturo, NDVI in barvni prostor (barva, nasičenost, intenziteta).



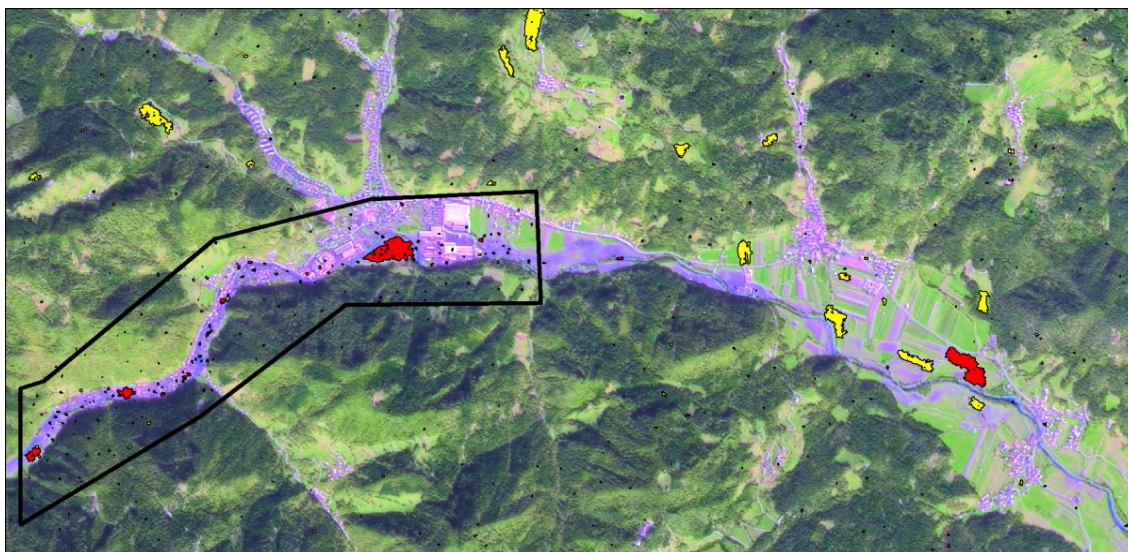
Slika 5: Učni vzorec z 255 točkami za učenje, razporejenih po celotnem obravnavanem območju na barvnem RGB večspektralnem posnetku SPOT, prikazanem s kombinacijo kanalov 4, 3 in 2. Rdeče točke (24) ležijo na poplavljenih in rumene (231) na nepoplavljenih tleh.

Figure 5: Sample of 255 training points lying on the whole treated area on multispectral image SPOT, shown with combination of 4, 3, 2 image channels. Red points (24) are lying in flooded and yellow points (231) in non-flooded areas.



Slika 6: Učni vzorec s 145 točkami za učenje, razporejenih na manjšem obravnavanem območju. Rdeče točke (50) ležijo na poplavljenih in rumene (95) na nepoplavljenih tleh.

Figure 6: Sample of 145 training points lying in the smaller part of the treated area. Red points (50) are flooded and yellow points (95) are lying in non-flooded areas.



Slika 7: Učni vzorec s segmenti. Učni vzorec sestavlja 255 segmentov, razporejenih po celotnem obravnavanem območju, in 143 segmentov, razporejenih na manjšem izseku (območje v okvirju).

Figure 7: Sample of 255 training segments lying on the whole treated area and 143 segments lying on the smaller part (framed area).

Preizkus klasifikacijskih algoritmov

Z namenom pridobitve modela, ki bi omogočal čim natančnejšo klasifikacijo poplavljenih površin, smo preizkusili delovanje algoritma odločitvenih dreves J48 (Quinlan, 1992; Witten in Frank, 2005), algoritma generiranja pravil JRip (Cohen, 1995) ter meta algoritma gozdov (Breiman, 1996) in naključnih gozdov (Breiman, 2001) (preglednica 2). Strojno učenje z vsemi preizkušenimi atributi je potekalo z učnim vzorcem točk, atributi o oblikovanosti površja pa so bili izpeljani iz DMV 0,5. Iz pridobljenih klasifikacijskih modelov so bile izdelane karte poplavljenih površin na celotnem območju. Za vsako od klasifikacij je izračunana ocena uspešnosti učenja z metodo 10-kratnega navzkrižnega preverjanj in ocena natančnosti klasifikacije, ki opisuje delež pravilno klasificiranih referenčnih/testnih točk, ki so bile razporejene tako na poplavljenih kot nepoplavljenih delih površja. Za klasifikacijo z najuspešnejšim algoritmom sta bili ocenjeni tudi izdelovalčeva in uporabnikova natančnost (Oštir, 2006). Prvo izračunamo kot delež pravilno klasificiranih referenčnih točk in drugo kot delež pravilno klasificiranih točk glede na celotno število točk, ki so bile v postopku klasifikacije uvrščene v določeno kategorijo, v našem primeru v kategorijo poplavljenih površin. V primeru poplavljenih površin uporabnikova natančnost torej pove, kolikšen del ozemlja, ki je bilo uvrščeno v razred poplavljenih površin, je bilo tudi v resnici poplavljenega. Preostanek površja namreč ni bil poplavljen in je zato njegova uvrstitev v razred poplavljenih površin napačna. Na drugi strani izdelovalčeva natančnost predstavlja podatek o deležu prepoznanih dejansko poplavljenih površin, z drugimi besedami določa verjetnost, da je bila površina, ki je klasificirana kot poplavljen, tudi v resnici pod vodo. Preostali delež poplavljenega površja je bil neprepoznan in uvrščen v razred nepoplavljenih površin.

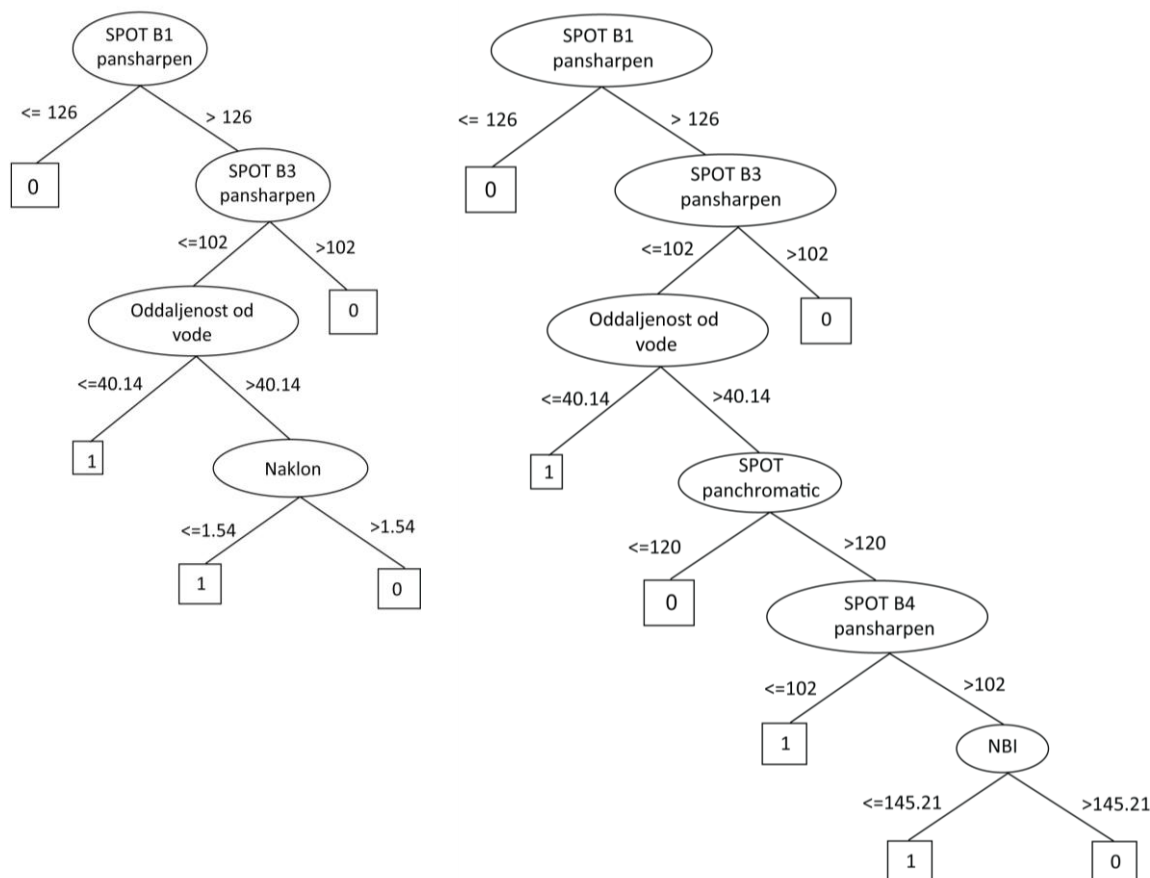
5.1.3 Rezultati

Predstavljeni so rezultati treh različnih preizkusov. Določili smo natančnost klasifikacije glede na ločljivost uporabljenega digitalnega modela višin, glede na obliko in velikost učnega vzorca sestavljenega iz točk oz. segmentov in glede na uporabljen algoritem strojnega učenja. Poplavljenе površine so nato tudi kartirane.

Rezultat učenja z DMV 0,5 in DMV 12,5

Rezultati učenja so pokazali, da višja ločljivost DMV izboljša natančnost klasifikacije poplavljenih površin na hribovitem območju Železnikov za 3 %. Natančnost zaznanih poplavljenih površin je bila 95 % v primeru uporabe DMV 0,5 in 92 % v primeru uporabe DMV 12,5. Razlog takšnega izboljšanja lahko določimo s primerjavo obeh modelov, ki sta bil pridobljena z uporabo DMV 0,5 (slika 8 - levo) in DMV 12,5 (slika 8 - desno). Oblika modelov je odločitveno drevo, ki s svojo strukturo zelo razumljivo predstavlja potek klasifikacije. Od zgoraj navzdol (od korenine do listov drevesa) si sledijo vozlišča dreves z atributi, ki usmerjajo klasifikacijo. Vozlišča so povezana z vejami dreves, ki določajo mejno vrednost posameznega atributa. Drevo se zaključi z listi, ki določajo pripadnost poplavljenim oz. nepoplavljenim površinam. Zgornja polovica odločitvenega drevesa je enaka tako ob upoštevanju DMV 0,5 kot DMV 12,5. V prvih treh vozliščih si sledijo atributi izostrenega (pansharpen) prvega (zelenega) kanala večspektralnega satelitskega posnetka SPOT, izostrenega tretjega (bližnjega IR) kanal večspektralnega satelitskega posnetka SPOT in oddaljenost od vodotokov. Območja z vrednostmi zelenega kanala 126 ali manj so že v začetku klasifikacije uvrščena med nepoplavljene površine. Ostala območja, ki imajo vrednost zelenega kanala višjo od 126, nadaljujejo postopek klasifikacije v vozlišče bližnjega IR kanala. Na tem mestu so med nepoplavljena območja uvrščeni tudi piksli z vrednostmi bližnjega IR kanal večjimi od 102. Sledi upoštevanje dodatnega pogoja oddaljenosti od vodotokov, kjer prvič dobimo tudi razred poplavljenih površin. Vsi piksli, katerih vrednost zelenega kanala je večja od 126, vrednost bližnjega IR kanala enaka oz. manjša od 102 in oddaljenost od vodotokov enaka ali manjša od 40 metrov, so uvrščeni v razred poplavljenih površin. V četrtem vozlišču se pojavi razlika med obema klasifikacijskima drevesoma. V modelu z upoštevanjem DMV 0,5 se na tem mestu pojavi še zadnje vozlišče z atributom naklona, ki piksle z naklonom večjim od 1,54 stopinj uvrsti v razred nepoplavljenih površin in piksle z naklonom manjšim ali enakim od 1,54 stopinj v razred poplavljenih površin. Pri upoštevanju DMV 12,5 se model namesto z vozliščem naklona nadaljuje s pankromatskim kanalom satelitskega posnetka SPOT, ki mu sledi vozlišče z atributom izostrenega (pansharpen) četrtega (kratkovalovnega IR) kanala večspektralnega satelitskega posnetka SPOT in vozlišče z indeksom NBI. Naklon nastopa le v modelu za klasifikacijo poplavljenih površin z uporabo DMV 0,5, saj sta ločljivost DMV 12,5 in njegova višinska natančnost, ki je za hribovita območja ocenjena na 3,8 m (Podobnikar in Mlinar, 2006), preslabi za natančno določitev naklona na lokalnih območjih. DMV 12,5 tako v nasprotju z DMV 0,5 ne omogoča natančne določitve struge Selške Sore in lokalih sprememb v oblikovanosti površja, kar zmanjšuje natančnost

določitve poplavljenih površin. Druga dva atributa, ki prav tako izvirata iz DMV, to sta višina in usmerjenost, nista vključena v nobeno od obeh klasifikacij, kar je posledica oblikovanosti površja na obravnavanem območju. Površje se spušča od zgornjega proti spodnjemu delu doline, kar vpliva na velike razlike v nadmorskih višinah, zaradi česar atribut nima vpliva na klasifikacijo. Podobno so poplavljeni tudi pobočja z različno usmerjenostjo površja (območja na obeh straneh struge), kar onemogoča prepoznavanje poplavljenih in nepoplavljenih površin glede na usmerjenost.



Slika 8: Odločitveni drevesi, pridobljeni z učenjem na digitalnem modelu višin DMV 0,5 (levo) in DMV 12,5 (desno). V obeh primerih je učenje potekalo z uporabo algoritma J48 in 145 učnih točk.

Figure 8: Decision trees built by 145 sampling points and algorithm J48. DEM 0,5 was used on the left and DEM 12,5 on the right example.

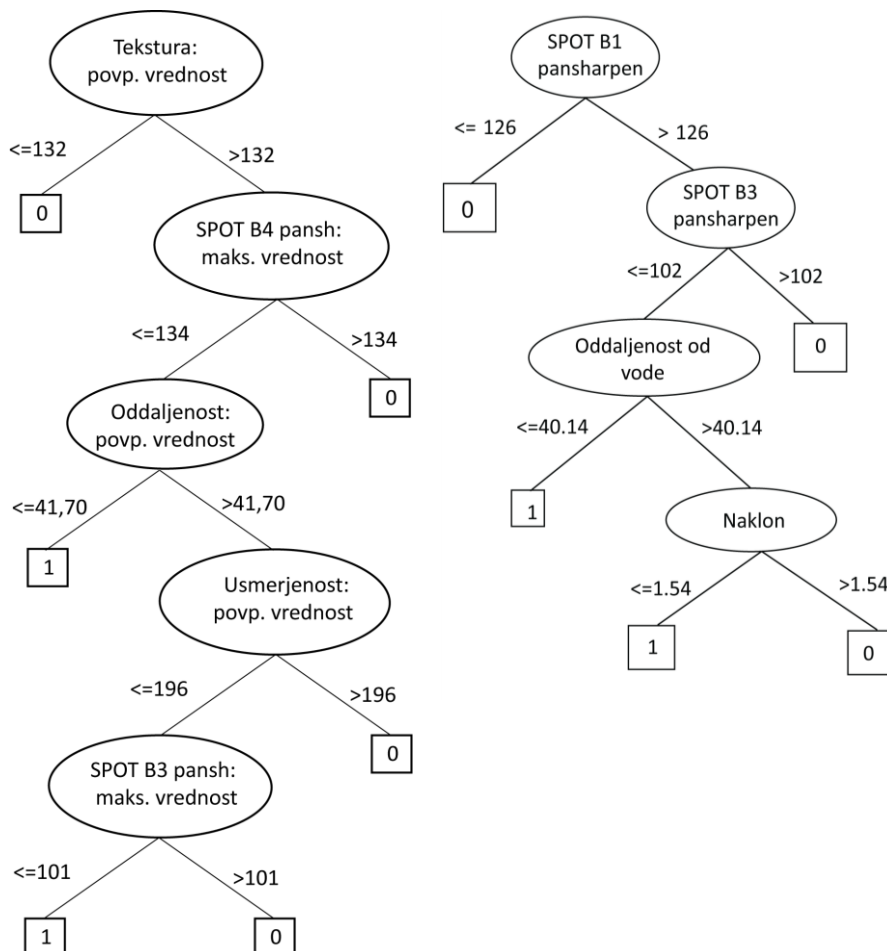
Natančnost DMV 12,5 se je izkazala za preslabo tudi v primeru hidravlične analize. Za določitev hidravličnih parametrov in izračun pretoka je nujna uporaba DMV 0,5, medtem ko DMV 12,5 ne omogoča določitve naklona struge z zadostno natančnostjo. Rezultati analize opravljene na DMV 0,5 so se ujemali z rezultati drugih avtorjev (Marchi in sod., 2010; Rusjan in sod., 2009). Vrhovi pretoka so znašali do okoli 300 m³/s, kar je omogočilo, da je vsa voda odtekla preko prizadetega območja že v nekaj urah. Opazovanje poplav iz satelitskih posnetkov je tako potekalo s pomočjo prepoznavanja

ostalnih znakov poplav (nanosi materiala, z vodo zalita nižje ležeča območja ipd.). Voda pa na poplavljenih površinah po treh dneh, ko je satelitski posnetek nastal, ni več vidna.

Rezultat učenja z različnimi učnimi vzorci

Natančnost klasifikacij z uporabo učnih vzorcev različne gostote je bila 81 % v primeru učnega vzorca z gostoto 11 točk / 100 ha in 86 % v primeru uporabe učnega vzorca z gostoto 58 točk / 100 ha. Z višjo gostoto učnih točk smo zagotovili bolj celovit opis obravnavanega območja z vsemi njegovimi raznolikostmi. Z večjo gostoto točk smo v učni vzorec vključili tudi posebnosti, kot so npr. točke, ki ležijo na višjih objektih, katerih vrhovi segajo nad gladino vode (npr. hiše, visoka trava, visoke poljščine in drevesa). Ti objekti sicer ležijo na poplavljenih tleh, vendar moramo točke, ležeče na teh območjih, označiti kot nepoplavljene, saj ima odboj EMV na teh površinah vse značilnosti odboja nepoplavljenega površja.

Druga primerjava učnega vzorca, sestavljenega iz točk oz. segmentov, je pokazala 95 % natančnost v primeru učnega vzorca s točkami in 91 % natančnost v primeru učnega vzorca s segmenti. Rezultat je v nasprotju z našimi pričakovanji pokazal, da klasifikacija s segmenti ne prinaša izboljšanja pri klasifikaciji poplavljenih površin. Odločitveno drevo, zgrajeno na osnovi točk (slika 9 - desno drevo), je enako drevesu, ki je bilo pridobljeno pri preizkusu gostote učnega vzorca (slika 8 - levo drevo). Na sliki 9 lahko vidimo, da pri obeh klasifikacijah nastopata vrednosti tretjega (bližnjega infrardečega) kanala izostrenega satelitskega posnetka SPOT in oddaljenost od vodotokov, ki so jima v obeh primerih dodeljene tudi zelo podobne mejne vrednosti. V klasifikacijo so vključeni še prvi (zeleni) kanal posnetka SPOT in naklon v primeru učenja s točkami oz. tekstura, četrti (srednji infrardeči) kanal posnetka SPOT in usmerjenost v primeru uporabe segmentov. Pri določevanju poplavljenih površin s segmenti sodelujejo poleg atributov, ki so določljivi samo za segmente (tekstura), tudi atributi, ki določajo spektralne lastnosti in oblikovanost reliefa, katere lahko določimo tudi točkam. Takšna zastopanost atributov, kjer tudi v modelu, pridobljenim z učenjem na segmentih, ne dominirajo samo lastnosti, ki so določljive izključno segmentom, pač pa se v klasifikacijo vključujejo tudi atributi, katerih primarne vrednosti se nanašajo na posamezen piksel / točko, za segment pa jim je mogoče določiti le statistične vrednosti (maksimum, minimum, povprečna vrednost), potrjuje, da je uporaba učnega vzorca točk za klasifikacijo poplavljenih površin na območju Železnikov primernejša od segmentov.



Slika 9: Odločitveni drevesi, pridobljeni z učenjem na učnem vzorcu segmentov (levo) oz. točk (desno). V obeh primerih je učenje potekalo na DMV 0,5 z uporabo algoritma J48.

Figure 9: Decision trees produced by training sample of segments (left) and training sample of points (right). The DEM 0,5 and algorithm J48 were used in both examples.

Ocena natančnosti klasifikacije z uporabo različnih algoritmov strojnega učenja

Preglednica 6 prikazuje uspešnost strojnega učenja in oceno natančnosti klasifikacije z uporabo štirih različnih algoritmov. Uspešnost učenja za vsako od metod strojnega učenja je določena z deležem pravilno klasificiranih točk za trening ob 10-kratnem navzkrižnem preverjanju. Za končno oceno natančnosti klasifikacije, ki določa pravilnosti karte poplavljenih površin smo uporabili 100 naključno generiranih testnih točk, ki so neodvisne od učne množice.

Preglednica 6: Uspešnost določitve poplav z uporabo različnih metod strojnega učenja. Med metodami za kombiniranje klasifikatorjev so prikazani rezultati z 10 drevesi.

Table 6: Accuracy of determination of flooded areas using different machine learning methods. Ensembles with ten trees are shown in the table.

Metoda:	Delež pravilno klasificiranih primerov	
	Učenje	Test
J48	88 %	95 %
JRip	86 %	85 %
Gozd J48 (10 dreves)	90 %	92 %
Naključni gozd C45 (10 dreves)	89 %	92 %

Odločitveno drevo je zelo praktična oblika za gradnjo modelov, zlasti v primerih z majhnim številom možnih razredov. V našem primeru gre za binarno klasifikacijo z dvema možnima razredoma: poplavljen in nepoplavljen površini. Najvišja natančnost klasifikacije, merjene s slučajno generiranimi testnimi točkami, je bila dobljena z odločitvenim drevesom J48. Skupna natančnost klasifikacije, ki opisuje delež pravilno klasificiranih referenčnih/testnih točk, razporejenih tako na poplavljenem kot nepoplavljenem delu površja, je bila 95 %. Iz matrike napak (preglednica 7) smo določili še izdelovalčevo in uporabnikovo natančnost. Višja izdelovalčeva in nižja uporabnikova natančnost kažeta na to, da so bile prepoznane skoraj vse poplavljen površine, da pa je bilo 9 % klasificiranih poplavljenimi površinami v resnici nepoplavljenih.

Preglednica 7: Matrika napak za oceno natančnosti klasifikacije na območju Železnikov.

Table 7: Confusion matrix to define classification accuracy in the study area of Železniki.

Razred	Poplavljen	Nepoplavljen	Klasificirane točke	Natančnost (%)	
				Izdelovalec	Uporabnik
Poplavljen	40	4	44	98	91
Nepoplavljen	1	55	56	93	98
Referenčne točke	41	59	100		

Klasifikacijska natančnost z obema uporabljenima metodama za kombiniranje klasifikatorjev (gozd in naključni gozd) je bila nekoliko nižja, obakrat 92 %. Najnižja natančnost pa je dobljena pri uporabi algoritma tvorjenja pravil JRip (85 %). Natančnost klasifikacij z uporabo množice dreves ni presegla natančnosti odločitvenega drevesa J48, kar je v nasprotju s pričakovanji. Najvišja dosežena natančnost metod za kombiniranje klasifikatorjev je bila 94 %, dobljena pri uporabi metode naključnih gozdov s 30 drevesi. Uporaba večjih množic ne prispeva k znatnemu povečanju natančnosti klasifikacije. Že z uporabo množice 10 dreves smo dosegli natančnost 92 %, ki se je pri 30 drevesih zvišala zgolj za 2 %, z nadaljnjim povečevanjem števila dreves pa se natančnost ne izboljšuje več (preglednica 8). Razlog za neučinkovitost klasifikacij z množico odločitvenih dreves je v količini podatkov, ki je bila

premajhna za uporabo večjega števila klasifikatorjev, saj se njihova učinkovitost lahko pokaže šele pri mnogo večji količini podatkov (Dietterich, 2000). V našem primeru pa je množica klasifikatorjev z 10 drevesi izčrpala večino novih informacij. Z nadaljnjim povečevanjem števila klasifikatorjev algoritmi strojnega učenja iščejo nove povezave med podatki, ki pa niso nujno pravilne. Posledica je prevelik vpliv atributov na potek klasifikacije, ki v resnici nimajo tako odločilnega pomena, kar povzroča šum v klasifikaciji in padanje natančnosti. Na območju Železnikov se je v primeru klasifikacije z večjo množico odločitvenih dreves povečal vpliv indeksov grajenega okolja (NBI, NDBI). Indeksa sta namenjena prepoznavanju pozidanih površin, zaradi podobnega odboja na pozidanih in poplavljenih površinah (prepoznavanje slednjih je potekalo s pomočjo posledic erozije in akumulacije materiala) pa sta povzročila tudi večje število zamenjav pozidanih površin s poplavljenimi. Množici odločitvenih dreves, pridobljeni z obema metodama za kombiniranje klasifikatorjev z 10 drevesi, sta prikazani v prilogi B za metodo gozdov in v prilogi C za metodo naključnih gozdov.

Preglednica 8: Natančnost določitve poplavljenih površin z uporabo metode naključnih gozdov (random forest) in različnim številom dreves.

Table 8: Accuracy of determination of flooded areas using random forest method with different number of trees.

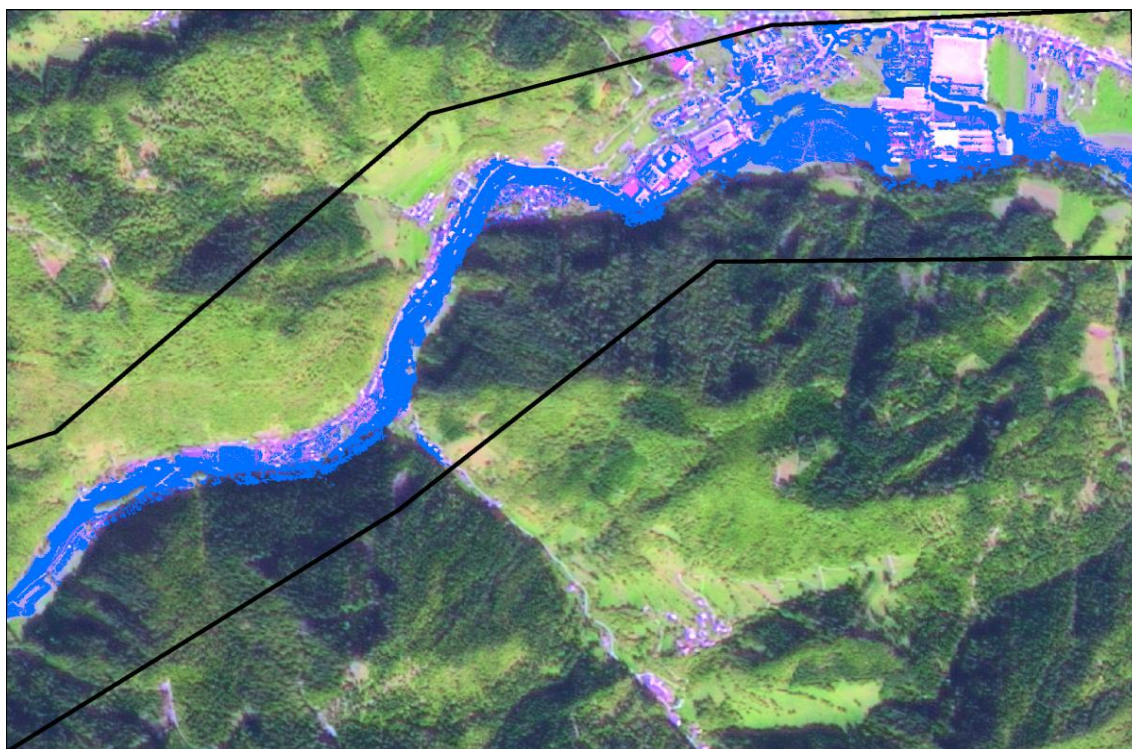
Velikost množice (število dreves)	Pravilno klasificirane učne točke	Pravilno klasificirane testne točke
5	90 %	86%
10	89 %	92%
20	90 %	92%
30	90 %	94%
40	89 %	93%
50	90 %	93%
60	91 %	93%

Pri tem je potrebno opozoriti, da je natančnost testa, ki je v večini primerov višja od natančnosti učenja (preglednica 6, 8), posledica izbire učnih točk. Te so bile razporejene tudi na za prepoznavanje poplavljenosti izrazito zahtevnih lokacijah, kot so npr. strehe stavb ali z grmičevjem porasla pobočja ob rečni strugi. V prvem primeru so točke označene kot nepoplavljene, medtem ko so v drugem primeru točke označene kot poplavljene, v kolikor je vodo pod grmičevjem mogoče prepoznati z vizualno interpretacijo satelitskega posnetka, in kot nepoplavljene, v kolikor je gostota rastja tolikšna, da povsem zastira pogled na tla. Učenje z zahtevno učno množico ima za posledico nižjo uspešnost učenja, ki pa je omogočilo natančnejšo klasifikacijo, ki je v večini primerov presegla natančnost učenja.

Kartiranje poplavljenih površin

Slika 10 prikazuje poplavljen območja prepoznana z uporabo klasifikacijskega modela, pridobljenega z algoritmom J48 na DMV 0,5. Model je bil uporabljen že ob preizkusih vpliva ločljivosti DMV in preizkusu vpliva oblike učnega vzorca na klasifikacijo poplavljenih površin. Atributi in njihove mejne

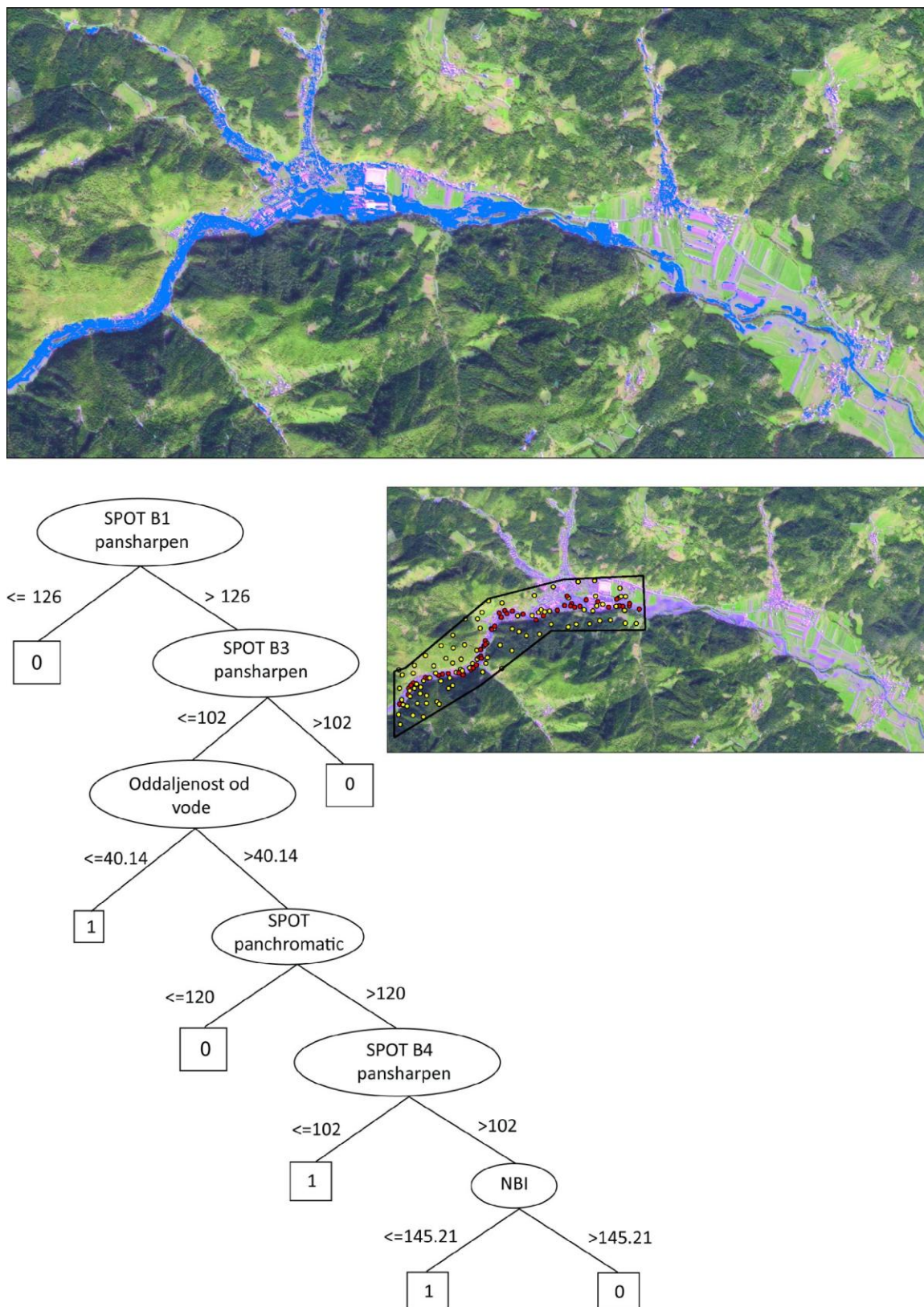
vrednosti so predstavljeni pri preizkusu ločljivosti DMV, njegova zgradba je vidna na slikah 8 (levo) in 9 (desno). Na karti so z modro barvo označena poplavljenega območja na večspektralnem satelitskem posnetku SPOT, prikazanem s kombinacijo četrtega (kratkovalovnega IR), tretjega (bližnjega IR) in drugega (rdečega) kanala. Posnetek je nastal tri dni po poplavih, ko je vsa voda že odtekla iz poplavljenega območja, zato smo si morali pri prepoznavanju poplavljeni površin iz satelitskih posnetkov pomagati z drugimi znaki, ki so bili vidni tudi tri dni po dogodku. Prizadeto območje je tudi tri dni po deževju ostalo zelo razmočeno, predvsem najnižja območja ob rečni strugi pa so bila tudi več dni po deževju se vedno delno poplavljenega. Na poplavljenih območjih so ostale tudi velike količine nanošenega material (blato, pesek), ki je na satelitskih posnetkih dobro vidno in nam pomaga pri določitvi poplavljenega ozemlja. Težavo pa povzroča podobnost odboja EMV na odloženem materialu, zemlji obdelanih kmetijskih zemljišč in pozidanih površinah. Podobnost je vidna tudi na večspektralnem satelitskem posnetku SPOT, prikazanem s kombinacijo 4., 3. in 2. kanala v rdeči, zeleni in modri barvi (sliki 10), kjer se vsa predstavljena območja, prikažejo v vijoličnih odtenkih, zaradi česar prihaja med temi površinami do pogostih zamenjav. S karte se vidi, da se največ napak pri klasifikaciji pojavlja na grajenih (stanovanjskih) in kmetijskih površinah (polja, njive).



Slika 10: Karta poplavljenih površin na območju Železnikov, pridobljena z algoritmom J48, 145 učnimi točkami in DMV 0,5. Poplavljene površine so predstavljene na večspektralnem satelitskem posnetku SPOT, prikazanem s kombinacijo kanalov kratkovalovne IR, bližnje IR in rdeče svetlobe.

Figure 10: Map of flooded areas in Železniki produced by algorithm J48, 145 sampling points and DEM 0,5. Flooded areas are represented on multispectral satellite image SPOT in combination of SWIR, NIR and red channels.

Model, ki ga pridobimo z izbranim algoritmom strojnega učenja, mora biti dovolj trden za njegovo uporabo na širšem ali drugem območju s podobnimi lastnostmi. Zato smo model, ki smo ga pridobili na manjšem območju za trening z učnim vzorcem sestavljenim iz 145 točk in uporabo DMV 12,5 (slika 8 - desno), uporabili za klasifikacijo večjega območja (slika 11). Test natančnosti klasifikacije s 125 testnimi točkami, razporejenimi po širšem območju, je pokazal 84 % pravilnost. Rezultat dokazuje tudi možnost ekstrapolacije modela z manjšega območja, kjer je bilo opravljeno učenje, za klasifikacijo na širšem območju. Ne moremo pa pričakovati, da bi isti model – brez modifikacij – lahko uporabili tudi na drugih, bolj oddaljenih območjih. Na bolj oddaljenih območjih s podobnimi značilnostmi reliefa obstaja sicer velika verjetnost podobne atributne sestave klasifikacijskega modela, vendar pa se bodo atributi razlikovali vsaj v njihovih mejnih vrednostih. V kolikor se nahajamo na območju drugačnimi reliefnimi lastnostmi, pa predvidevamo, da se bodo razlike pojavljale tudi v sestavi atributov, ki bodo vključeni v model za klasifikacijo. V kolikor bi želeli sestaviti model, ki bi veljal tudi na bolj oddaljenih lokacijah, bi bilo potrebno upoštevati samo attribute, katerih vrednosti so neodvisne od reliefa. To so zlasti spektralne lastnosti ter lastnosti teksture in geometrije objektov na površju, vendar pa bi neupoštevanje reliefnih značilnosti močno poslabšalo natančnost določevanja poplavljenih površin. Smiselna bi bila torej izdelava večjega števila modelov za klasifikacijo poplav na različnih območjih, s katerimi bi pokrili celotno ozemlje, na katerem obstaja možnost poplav, sledila pa bi primerjava teh modelov in iskanje zakonitosti v njih.



Slika 11: Karta poplavljenih površin celotnega obravnavanega območja, pridobljena z algoritmom J48 (zgoraj). Spodaj sta prikazana odločitveno drevo in uporabljeni učni vzorec s 145 točkami, razporejenimi na manjšem izseku.

Figure 11: Map of flooded areas in the whole treated area produced by algorithm J48 (above). Decision tree and sample of 145 training points arranged on the smaller part are shown below.

5.1.4 Razprava

Učenje na različnih oblikah učnega vzorca (točke oz. segmenti) in ob uporabi digitalnega modela višin dveh različnih ločljivosti (DMV 12,5 in DMV 0,5) je pokazalo uspešnejšo klasifikacijo pri učnem vzorcu točk in DMV 0,5. Natančnost klasifikacije se je zvišala z 91 % na 95 % (za 4 %) v primeru uporabe točk namesto segmentov in z 92 % na 95 % (za 3 %) v primeru uporabe DMV 0,5 namesto DMV 12,5. Klasifikaciji s segmenti in DMV 12,5 nudita nekoliko nižjo, a še vedno zadovoljivo natančnost. Razlog višje natančnosti klasifikacije z učnim vzorcem točk, v primerjavi s segmenti, je v postopku segmentacije satelitskega posnetka, ki zaradi velike spektralne podobnosti poplavljenih površin s pozidanimi območji in njivami ne omogoča pravilne izločitve poplavljenih območij od nepoplavljenih. Na drugi strani je višja natančnost klasifikacije ob uporabi DMV 0,5 posledica nizke ločljivosti DMV 12,5, ki ne omogoča natančne določitve naklona na lokalnem območju, ampak predstavlja povprečno vrednost širšega območja, kar pa ne pripomore k natančnejšemu prepoznavanju poplavljenih površin na lokalnem območju. Pri tem je potrebno upoštevati, da zgoraj navedeni oceni natančnosti vključujeta tako površine, ki so bile nepravilno uvrščene v razred poplavljenih površin, kot površine, ki so bile poplavljene, pa jih klasifikacija ni prepoznala za poplavljeno območje. Na območju Železnikov prevladuje prekomerna zaznava poplavljenih površin na urbanih in kmetijskih površinah, medtem ko se primeri premajhne zaznave poplavljenih površin pojavljajo le izjemoma.

Pri klasifikacijah z učnim vzorcem točk oz. segmentov se je pokazala podobnost v uporabljenih atributih. V obeh primerih sta bila uporabljena bližnji infrardeči kanal posnetka SPOT in oddaljenost od vodotokov. Preostali atributi pa so opisovali spektralne lastnosti, lastnosti reliefa in v primeru uporabe segmentov tudi atribut teksture. Tudi učenje na digitalnem modelu reliefa dveh različnih ločljivosti je imelo za posledico dve zelo podobni odločitveni drevesi, ki se razlikujeta le v zaključnem delu drevesa. Te podobnosti kažejo na jasno določljive lastnosti, s katerimi se poplavljene površine razlikujejo od nepoplavljenih. Preglednica 9 prikazuje attribute, ki so bili kot najpomembnejši prepoznani v postopku učenja s štirimi različnimi algoritmi. Učenje je v vseh prikazanih primerih potekalo na učnem vzorcu točk in digitalnem modelu višin DMV 0,5. Atributi, ki so vključeni pri klasifikacijah z enim drevesom (v primeru algoritmov J48 in JRip), so oddaljenost od vodotokov, naklon in bližnji infrardeči kanal posnetka SPOT. Pri klasifikaciji je upoštevan še zeleni kanal v modelu, pridobljenim z algoritmom J48, in indeks NDVI v modelu, pridobljenim z algoritmom JRip. Vsi ti atributi igrajo vodilno vlogo tudi pri klasifikacijah, ki so potekale s kombiniranjem množice odločitvenih dreves, kjer pri klasifikaciji vsaj enkrat sodelujejo tudi vsi ostali atributi, vključeni v učenje. Nobeden od uporabljenih atributov ni imel negativnega vpliva na pravilnost klasifikacije, kar pomeni, da so lahko v postopek učenja vključeni vsi atributi, algoritem pa nato prepozna najpomembnejše izmed njih in iz njih sestavi model za klasifikacijo poplavljenih površin. Med atributi, ki so se pojavili vsaj v dveh drevesih celotne množice, so bili še atribut usmerjenosti, izpeljan

iz digitalnega modela višin, indeksa NBI in NDBI, pankromatski in srednji infrardeči kanal posnetka SPOT (2,5 m) ter bližnji infrardeči kanal posnetka SPOT ločljivosti 10 m.

Preglednica 9: Prikaz atributov, ki sodelujejo pri klasifikacijah poplavljenih površin na območju Železnikov z učenjem na učnem vzorcu točk in različnimi algoritmi. Vrednosti v preglednici predstavljajo število modelov, v katerih se pojavi posamezen atribut. Večje število pojavljanj pomeni večji vpliv atributa na klasifikacijo.

Table 9: Attributes included into the model for classification of flooded areas in the area of Železniki. The learning process was performed using sample of points and different algorithms. Values in the table present the number of models where the attributes are taking part. Larger values of the attributes show higher impact on the classification process.

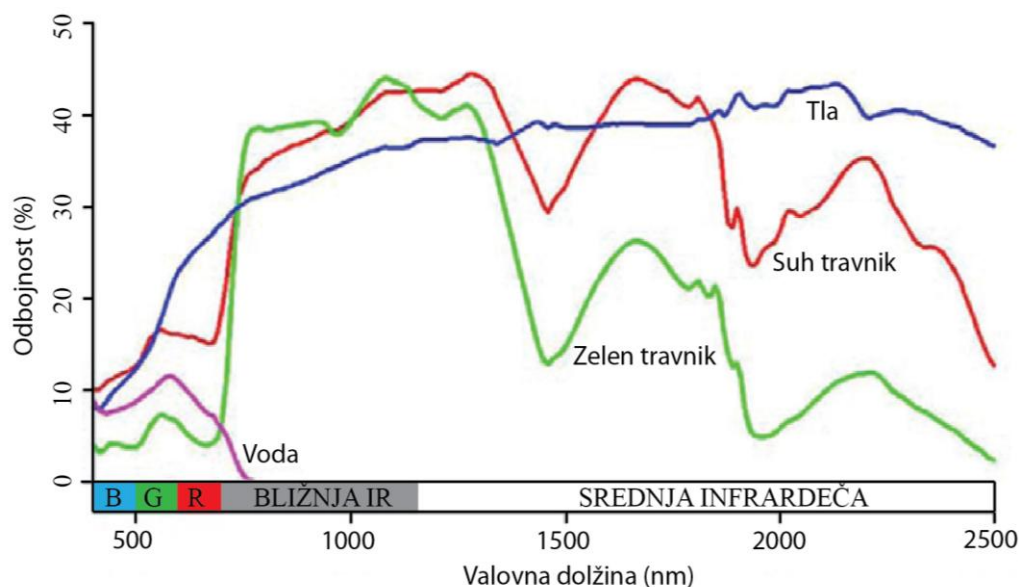
Atribut	Gozd*	Naključni gozd*	J48	JRip
B1 – zelen (10 m)	0	1	0	0
B2 – rdeč (10 m)	0	3	0	0
B3 – bližnji IR (10 m)	3	3	0	0
B4 – srednji IR (10 m)	2	1	0	0
B1 – zelen (2,5 m)	3	4	1	0
B2 – rdeč (2,5 m)	1	5	0	0
B3 – bližnji IR (2,5 m)	2	3	1	1
B4 – srednji IR (2,5 m)	7	4	0	0
Pan (2,5 m)	4	2	0	0
NDVI	6	5	0	1
NBI	2	5	0	0
NDBI	2	3	0	0
Višina	1	5	0	0
Naklon	4	2	1	1
Usmerjenost	3	2	0	0
Oddaljenost	7	3	1	1

*Metodi gozdov in naključnih gozdov sta potekali iz izgradnjo 10 dreves.

Bližnji infrardeči in zeleni kanal satelitskega posnetka SPOT ločljivosti 2,5 m, oddaljenost od vodotokov, naklon in indeks NDVI so atributi, ki se najpogosteje pojavljajo v klasifikacijskih modelih in so torej potrebni za določevanje poplavljenih površin v hribovitih predelih.

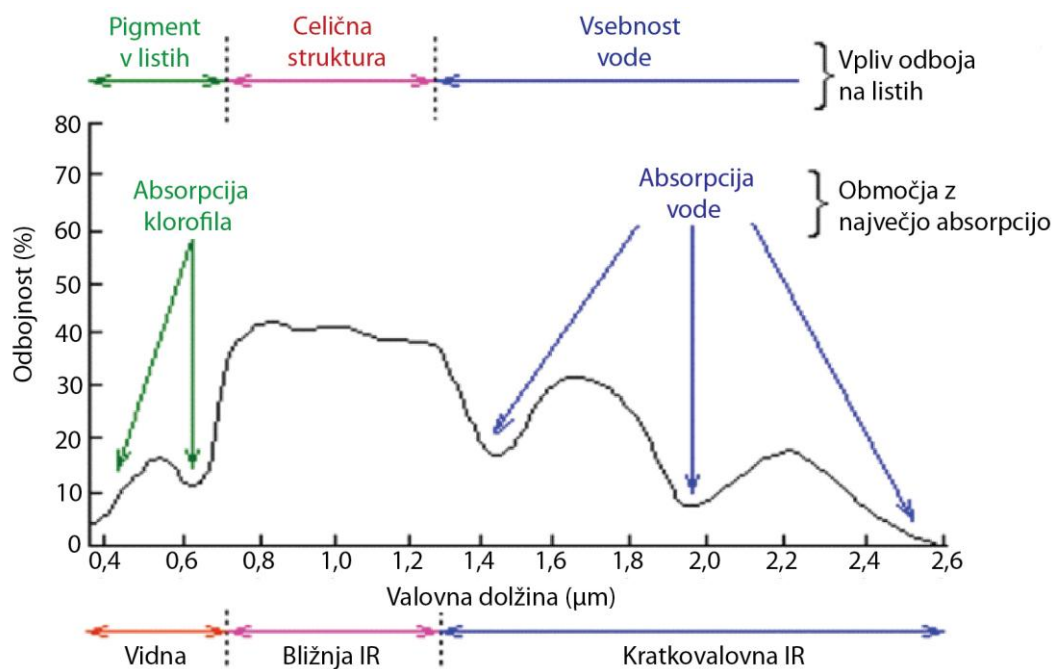
Voda v primerjavi z ostalimi površinami absorbira večji delež bližnje infrardeče svetlobe, zaradi česar so poplavljene površine v tem delu spektra videti temnejše. Večji del zelene svetlobe pa se od vodnih površin odbije in poplavljene površine se v tem delu spektra pokažejo svetlejše (slika 12). Indeks NDVI je bil vključen v vse klasifikacijske modele, z izjemo modela, ki je sicer zagotavljal najvišjo natančnost (pridobljen z algoritmom J48). Indeks NDVI, ki je sicer značilen kot dober indikator zelenih rastlin, ni sodeloval pri klasifikaciji s tem modelom zaradi zgodnjega jesenskega obdobja, ko rastline že začenjajo izgubljati klorofil in se zato poveča tudi odboj v rdečem delu spektra (slika 13). Pomemben atribut za razlikovanje poplavljenih in nepoplavljenih površin je bil tudi naklon, saj so območja z nižjimi nakloni ob rečni strugi poplavljen, medtem ko ostanejo bolj strma pobočja, ki so hkrati bolj oddaljena od rečne struge, nepoplavljen. Oddaljenost od vodotokov je tako naslednji pomembni dejavnik za določitev poplavljenih površin. Za razliko od naklonov nadmorske višine

nimajo pomembnejšega vpliva na klasifikacijo poplavljenosti, saj se višine poplavljenih površin v zgornjem in spodnjem delu opazovane doline med sabo zelo razlikujejo.



Slika 12: Spektralni odboj na različnih pokrovnostih tal (voda, zelena travnik, suh travnik, njiva) glede na valovno dolžino EMV (Ashraf in sod., 2011).

Figure 12: Reflectance spectrum of different land cover types (water, green grass, dry grass, soil) at different wavelengths (Ashraf in sod., 2011).



Slika 13: Spektralni odboj na vegetaciji. Lepo se vidi vpliv klorofila, celične strukture in vsebnosti vode v listih na odbojnost v različnih območjih vidne in IR svetlobe (Ashraf in sod., 2011).

Figure 13: Reflectance spectrum of vegetation. Influence of chlorophyll, cell structure and water content on the reflectance is seen (Ashraf in sod., 2011).

Z uporabo algoritma J48 se je izboljšala natančnost prepoznavanja poplavljenih površin na nekaterih območjih, kjer je določevanje poplavljenosti še posebej zahtevno. Zlasti presenetljivo je uspešno zaznavanje robu reke pod gozdnim drevjem (slika 14). Poplavljene površine prekrite z gozdom dosegajo visoke vrednosti v zelenem delu elektromagnetnega spektra ($B1 > 126$) in nizke vrednosti v bližnjem IR delu elektromagnetnega spektra ($B3 \leq 102$), kar omogoča njihovo prepoznavanje. Ostala dva pogoja sta dokaj predvidljiva, in sicer oddaljenost od vod ne sme presegati 40 m in naklon ne sme biti večji od $1,5^\circ$ (odločitveno drevo na sliki 10).



Slika 14: Poplavljene površine pod gozdom. S puščico so označena mesta uspešno prepoznanega robu poplavljenega območja na mestih prekritih z gozdom (prikaz na večspektralnem satelitskem posnetku SPOT - levo oz. letalskem posnetku DOF - desno).

Figure 14: Flooding under the forest. Red arrows sign the places where the borders of the flooded areas were detected successfully under the forest (presented on the multispectral SPOT image - left and digital orthophoto DOF - right).

Problematično pa je prepoznavanje poplav na pozidanih površinah in na območjih namenjenih poljedelstvu, kjer težave niso bile povsem rešljive. Poplav med strnjeno stanovanjsko gradnjo s satelitskih posnetkov SPOT z ločljivostjo 2,5 m ne moremo natančno prepoznati. Ločljivost je pregroba in zato na teh površinah prihaja do zamenjav med poplavljenimi in pozidanimi površinami, ki so še pogostejše v primerih stavb s temnejšo strešno kritino. Na drugi strani se večji gospodarski objekti zelo lepo ločijo od poplavljene okolice (slika 15).



Slika 15: Poplavljene površine na pozidanih območjih. Manjših stavb in poplavljenih površin med njimi ni mogoče razlikovati (rdeča puščica). Druge/večje stavbe se enostavno ločijo od poplavljenih površin.

Figure 15: Flooding in urban areas. Smaller buildings and flooded areas could not be distinguished completely. Other/larger buildings are distinguished easily from flooded areas.

Pogosto so kot poplavljenaa zaznana tudi polja in njive, s katerih je bil pridelek že pospravljen. Razlog je v odbitem elektromagnetnem valovanju, ki je na prsti kmetijskih površin lahko zelo podoben odboju na poplavljenih površinah (slika 16). Do podobnih vrednosti na kmetijskih in poplavljenih površinah prihaja predvsem v rdečem in srednjem infrardečem kanalu, zato je za uspešno prepoznavanje posameznih površin pomembna vključenost različnih atributov v proces učenja.



Slika 16: Poplavljenaa površine na kmetijskih tleh. Do zamenjav kmetijskih površin s poplavljenimi prihaja predvsem na območjih njiv in polj, s katerih je bil pridelek že pospravljen (območja z orno zemljo).

Figure 16: Flooding in agricultural areas. Conflicts between agricultural and flooded areas are happening. Especially cultivated areas of fields are often misclassified as flooded.

Zaključimo lahko, da smo s postopkom strojnega učenja povečali natančnost določitve poplavljenih površin. Še vedno pa prihaja zaradi podobnosti oddanega valovanja do zamenjav med kmetijskimi in poplavljenimi površinami. Natančnejše zaznavanje poplav na pozidanih območjih bi lahko dosegli z uporabo posnetkov višje prostorske ločljivosti, ki pa za našo študijo niso bili na voljo.

Z uporabo tehnik strojnega učenja, ki pomagajo pri določevanju najpomembnejših atributov satelitskih posnetkov in drugih podatkov, se je zvišala tudi stopnja samodejnosti postopka prepoznavanja poplavljenih površin. Določeni so bili najvplivnejši atributi in najuspešnejši algoritmi za učenje. S tem je čas, ki je potreben za izdelavo končne karte poplavljenih površin, krajši, kar je v primerih naravnih nesreč ključnega pomena.

5.2 Ljubljansko barje - primer kraškega tipa poplav

5.2.1 Podatki

Določitev poplavljenih površin po postopku strojnega učenja na območju Ljubljanskega barja je potekala na optičnih večspektralnih satelitskih posnetkih RapidEye. Poleg njih so bili v doktorski disertaciji uporabljeni še radarski satelitski posnetki Envisat in Radarsat ter aeroposnetki v infrardečem delu spektra (IR DOF) (preglednica 10), ki pa niso bili vključeni v postopek strojnega učenja. Radarski posnetki so bili uporabljeni za spremljanje poplav v prvih sedmih dneh po koncu obilnih padavin. Posnetek Envisat je nastal takoj po prenehanju obilnih padavin, 19.9.2010, in so mu

sledili še trije posnetki Radarsat z dni 23.9., 26.9. ter 2.10 (Veljanovski in sod., 2011a; Veljanovski in sod., 2011b).

Aeroposnetki so bili uporabljeni le za pomoč pri določitvi atributa poplavljenosti učnim primerom (točkam oz. segmentom) in testnim točkam. Atribut poplavljenosti določa ali točka/segment pripada poplavljenim oz. nepoplavljenim površinam.

Poleg optičnih posnetkov RapidEye smo imeli za obravnavano območje na voljo še posnetke Landsat, vendar pa so njihovo uporabo preprečevali oblaki, ki so prekrivali območje Ljubljanskega barja. Poleg tega prostorska ločljivost posnetkov Landsat znaša 30 m, kar omogoča predvsem izdelavo preglednih kart, ne pa tudi detajlnega prepoznavanja poplavljenih površin na manjših območjih.

Preglednica 10: Uporabljeni satelitski in letalski posnetki za opazovanje poplavljenih površin na območju Ljubljanskega barja 2010.

Table 10: A list of satellite images that were used for observation of the floods in the Ljubljana Moor area.

Satelit	Ločljivost (m)	Datum	Vir	Vrsta	Uporaba
Landsat	30	21.9.2010	USGS	Optični	Pregledne karte
RapidEye	6,5	23.9.2010	RapidEye	Optični	Strojno učenje
Envisat	12,5	19.9.2010	ESA	Radarski	Časovna vrsta
Radarsat	12,5	23.9.2010 26.9.2010 2.10.2010	CSA	Radarski	Časovna vrsta
DOF IR	0,2	20.9.2010	ARSO	Optični	Atribut poplavljenosti

V postopek strojnega učenja sta bila vključena še digitalni model višin in sloj hidrografije – omrežje vodotokov. V primeru digitalnega modela višin smo imeli na razpolago izključno DMV ločljivosti 12,5 m. Po poplavah 2010 se je pričelo izvajati tudi snemanje celotnega ozemlja Slovenije z lidarjem (GURS, 2011), kar bo omogočilo izdelavo natančnejšega DMV, vendar podatki v času raziskave še niso bili na voljo. Uporabljeni sloj hidrografije je omogočal določitev vodotokov, vodnih kanalov in ostalih vodnih poti, ki jih potrebujemo za izračun oddaljenosti od vodnih teles. Atributi, ki smo jih uporabili za učenje na Ljubljanskem barju, so tako bili: barvni kanali satelitskega posnetka RapidEye, indeks NDVI, višina, naklon, usmerjenost, ukrivljenost ter oddaljenost od vodotokov (preglednica 11).

Preglednica 11: Podatkovni sloji in njihovi atributi, uporabljeni za prepoznavanje poplavljenih površin na območju Ljubljanskega barja. Pri posameznem kanalu posnetka RapidEye so v oklepaju podana tudi spektralna območja zaznanega elektromagnetnega valovanja v nanometrih (nm).

Table 11: Data and their attributes used for determination of flooded areas in the area of Ljubljana Moor. Wavelengths of each spectral band of RapidEye sensors are given in brackets in the second column.

Poplave Ljubljansko barje (2010): 11 atributov	
1.) RapidEye	
večspektralni posnetek (6,5 m)	B1 (modra: 440 – 510 nm) B2 (zelena: 520 – 590 nm) B3 (rdeča: 630 – 685 nm) B4 (rob rdeče: 690 – 730 nm) B5 (bližnja IR: 760 – 850 nm)
indeks	NDVI
2.) DMV 12.5	Višina Naklon Usmerjenost Ukrivljenost
3.) HIDROLOGIJA (omrežje rek)	Oddaljenost od vodotokov

5.2.2 Metode dela

Podobno kot v primeru Železnikov je bil tudi na območju Ljubljanskega barja z namenom čim pravilnejše določitve poplavljenih površin opravljen preizkus strojnega učenja z različnima oblikama učnih primerov in različnimi algoritmi (preglednica 12).

Preglednica 12: Strojno učenje na območju Ljubljanskega barja. Preglednica prikazuje različne kombinacije strojnega učenja, ki so bile preizkušene glede na obliko učnih primerov in vrsto uporabljenih algoritmov.

Table 12: Machine learning in study area of Ljubljana Moor. Table presents different combinations training samples and machine learning algorithms which were used for machine learning process.

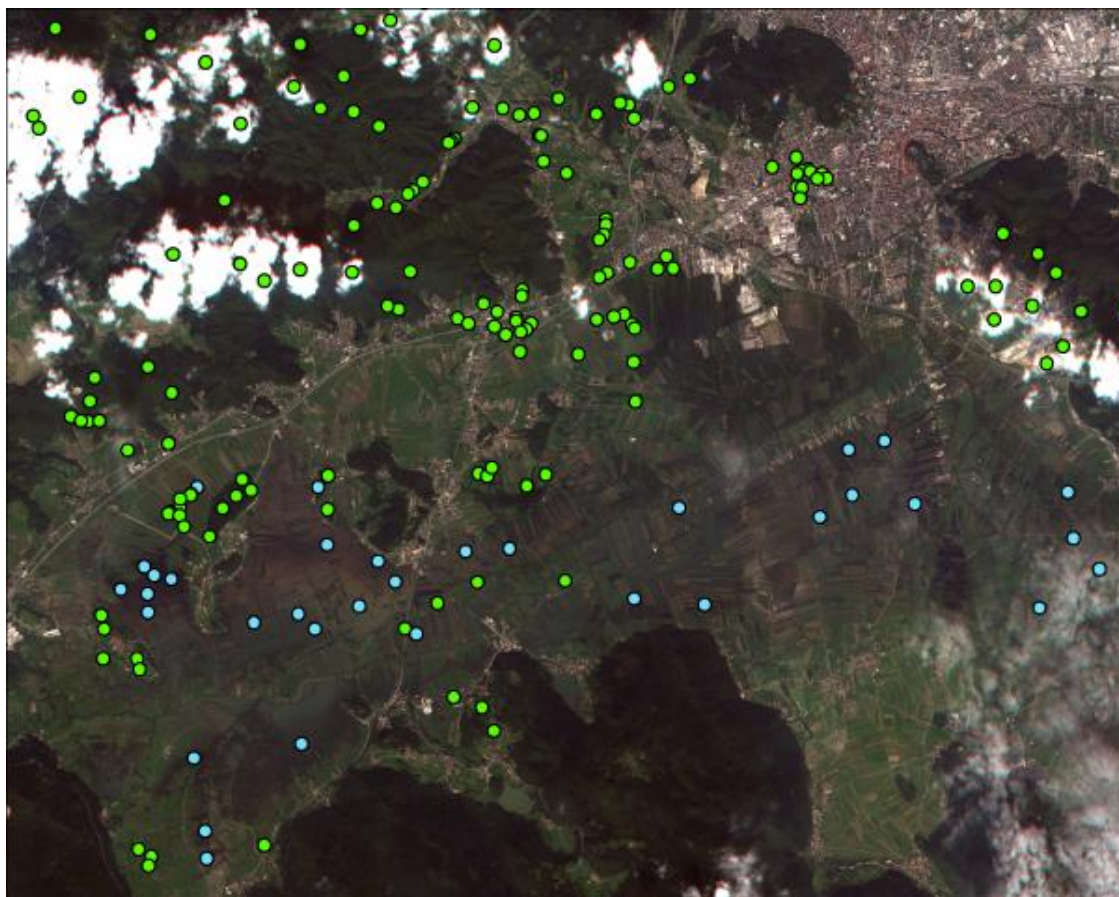
Vrsta preizkusa	Ločljivost DMV	Št. učnih primerov / velikost območja	Oblika učnega vzorca	Algoritem
Preizkus vpliva učnega vzorca - oblika	DMV 12,5	201 točk/segmentov (58 točk / ha)	<i>točke in segmenti</i>	J48
Preizkus klasifikacijskih algoritmov	DMV 12,5	201 točk/segmentov (58 točk / ha)	točke	<i>J48, JRip, gozd in naključni gozd</i>
			segmenti	<i>J48, podporni vektorji in najbližji sosed</i>

Iz modelov, pridobljenih s postopkom strojnega učenja, smo opravili posamezne klasifikacije in določili njihove natančnosti. Za razliko od primera poplav v Železnikih, pa klasifikacija na

Ljubljanskem barju ni vključevala preizkusa vpliva DMV in gostote učnega vzorca na natančnost klasifikacije. Na območju Ljubljanskega barja smo imeli na razpolago samo DMV 12,5, zaradi česar preizkus uporabe različnih ločljivosti DMV ni bil mogoč. Prav tako ni bil ponovljen preizkus učenja z različno gostoto učnega vzorca. Ugotovitve glede razporeditve učnih primerov, dobljene na primeru Železnikov, pa so bile upoštevane tudi pri izbiri učnega vzorca na območju Ljubljanskega barja.

Preizkus vpliva učnega vzorca

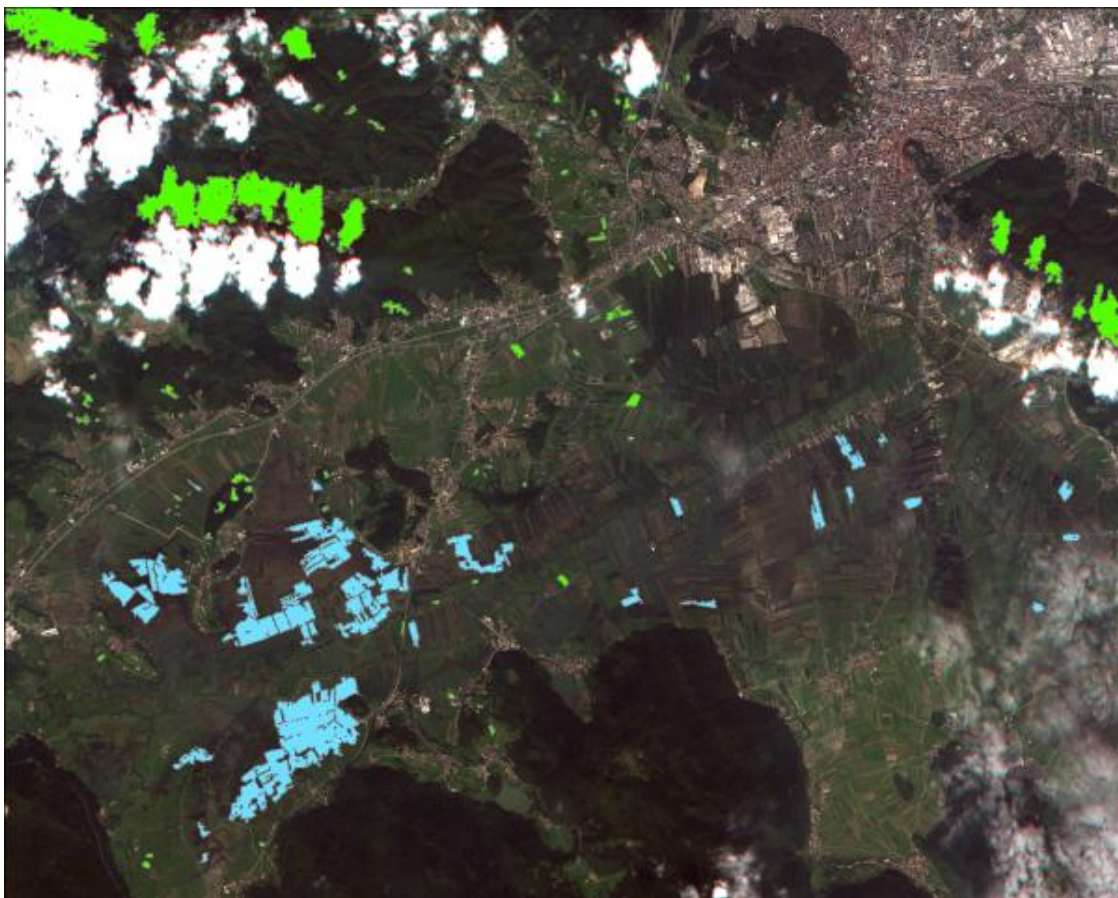
Učni vzorec točk je bil sestavljen iz 201 učnega primera, ki so bili dobljeni z naključno porazdelitvijo. Edini pogoj, ki smo ga zahtevali, je bil, da so bili učni primeri razporejeni po različnih vrstah rabe tal in da je bil delež učnih primerov na posamezni vrsti rabe tal podoben. Tako smo zagotovili, da so bili učni primeri enakomerno razporejeni na območjih gozdov, na njivah, oblakih in njihovih sencah, travnikih, urbanih površinah in vodnih oz. poplavljenih območjih. Najprej smo pripravili učni vzorec sestavljen iz točk (slika 17), ki smo jih nato nadomestili z enakim številom segmentov, razporejenimi na istih lokacijah kot točke (slika 18). Učne točke opisujejo obravnavano območje z 11 atributi (preglednica 11), ki predstavljajo odboj EMV v petih spektralnih kanalih satelitskega posnetka RapidEye, indeks NDVI, nadmorsko višino, naklon, usmerjenost, ukrivljenost in oddaljenost od vodotokov. Segmenti so bili pridobljeni v postopku segmentacije z modulom Feature Extraction v programu ENVI in vsebujejo 42 atributov. Vsak segment je predstavljen s 14 atributi, ki opisujejo geometrijske lastnosti segmentov, štirimi spektralnimi atributi za vsak posamezen kanal večspektralnega posnetka, štirimi teksturnimi atributi, normirano vrednostjo razmerja dveh kanalov večspektralnega posnetka, ki v našem primeru predstavlja vrednost NDVI in barvnim prostorom HSI, predstavljenim z barvo, nasičenostjo in intenziteto. Primerjava uspešnosti učenja s točkami in segmenti je bila analizirana z uporabo algoritma J48, ki se je med preizkušenimi algoritmi izkazal kot najučinkovitejši.



Slika 17: Vzorec z 201 učno točko, ki so razporejene na obravnavanem območju Ljubljanskega barja na barvnem RGB posnetku RapidEye, prikazanim s kombinacijo kanalov 3, 2, 1. Modre točke (34) ležijo na poplavljenih in zelene (167) na nepoplavljenih tleh.

Figure 17: Sample of 201 training points lying on the treated area of Ljubljana Moor on multispectral image RapidEye, shown with combination of 3, 2, 1 image channels. Blue points (34) are lying on flooded and green points (167) on unflooded areas.

Tako kot v primeru Železnikov, smo tudi na območju Ljubljanskega barja želeli pridobiti čim manjše segmente, zato smo izdelali podrobno segmentacijo satelitskega posnetka RapidEye s stopnjo merila 20 in stopnjo združevanja segmentov 20. S tem smo zagotovili čim podrobnejši opis območja s segmenti. Segmentom namreč ne moremo določiti absolutne vrednosti atributa, npr. odboja EMV v izbranem kanalu, kot je to v primeru točke. Vrednosti atributov na območju segmentov, npr. vrednosti odboja EMV v izbranih kanalih, lahko za segmente podamo le v obliki statističnih vrednosti, kot so minimalna, maksimalna in povprečna vrednost ter standardni odklon. Segmentom pa lahko poleg atributov, ki jih določamo točkam, dodamo še attribute, ki so določljivi samo za skupino pikslov oz. območje. Med njimi so najpogostejše uporabljeni tekstura in atributi, ki opisujejo geometrijske lastnosti segmentov, npr. površina, obseg in dolžina glavne osi segmenta. Največji atributi na sliki 18 pripadajo sencam oblakov in poplavljenimi površinam. Ta območja nam ni uspelo razdeliti na manjše segmente, saj jih v celoto povezujejo zelo podobne vrednosti spektralnih atributov in podobna geometrija.



Slika 18: Vzorec z 201 ucnim segmentom, ki so razporejeni na obravnavanem območju Ljubljanskega barja na barvnem RGB posnetku RapidEye, prikazanim s kombinacijo kanalov 3, 2, 1. Modro obarvani segmenti (34) ležijo na poplavljenih in zeleni (167) na nepoplavljenih tleh.

Figure 18: Sample of 201 training segments lying on the treated area of Ljubljana Moor on multispectral image RapidEye, shown with combination of 3, 2, 1 image channels. Blue segments (34) are lying on flooded and green segments (167) on unflooded areas.

Na obravnavanem območju smo se srečevali z različno obarvanostjo vode na poplavljenih površinah. Pri izbiri ucnega vzorca smo zato morali paziti, da so v vzorec vključeni primeri različnih barvnih odtenkov, ki se pojavljajo na poplavljenih površinah, saj v nasprotnem primeru ne bi bile zaznane poplavljene površine v celoti. Različni barvni odtenki vode so posledica primesi v vodi, različnih globin, različnih lastnosti tal in senc oblakov. Te lastnosti se jasno izražajo v vidnem delu spektra. Najpogosteje gre za kombinacijo modre, rjave in zelene barve, izražene v različnih odtenkih (slika 19).



Slika 19: Radiometrična nehomogenost poplavljenih površin na barvnem satelitskem posnetku RapidEye. Poleg značilne modrozeleno barve so poplavljene površine obarvane tudi s temnejšimi rjavimi in črnimi barvnimi odtenki.

Figure 19: Spectral inhomogeneity of flooded areas on multispectral RapidEye image. Beside usual blue and green colours, flooded areas are coloured also in brown and black.

Preizkus klasifikacijskih algoritmov

Za preizkus klasifikacijskih algoritmov strojnega učenja sta bila uporabljena specializirana programa za strojno učenje Weka in Clus, izvedena pa je bila tudi klasifikacija s programom ENVI. Klasifikacijski algoritmi v programih Weka in Clus so bili preizkušeni z učnim vzorcem točk in klasifikacijski algoritmi programa ENVI z učnim vzorcem segmentov. Izjema je bil algoritem J48 uporabljen v programu Weka, ki je bil preizkušen za določitev vpliva oblike učnega vzorca na natančnost klasifikacije in je bil uporabljen tako za učenje s točkami kot segmenti. Postopek klasifikacije s strojnim učenjem v programih Weka in Clus je bil podrobno predstavljen že v poglavju 3.1. V programu ENVI pa so bila poplavljena območja določena z uporabo modula Feature Extraction. Postopek klasifikacije z uporabo tega modula je zelo podoben postopku klasifikacije s pomočjo strojnega učenja v programih Weka in Clus, saj se prične s pripravo podatkov in iskanjem zelenih objektov s postopkom segmentacije. Nato med najdenimi objekti izberemo primere za sestavo učnega vzorca in vrsto algoritma za izvedbo nadzorovane klasifikacije. Uporabili smo oba algoritma za učenje klasifikacije, ki ju modul Feature Extraction ponuja: algoritem najbližjega soseda (nearest neighbor) (Roussopoulos in sod., 1995) in algoritem podpornih vektorjev (support vector machine) (Scholkopf, 1997; Hsu in sod., 2007). Izgradnja preostalih klasifikacijskih modelov s programoma Weka in Clus je potekala z algoritmi za izgradnjo odločitvenih dreves J48 (Quinlan, 1992; Witten in Frank, 2005), za

izgradnjo odločitvenih pravil JRip (Cohen, 1995) ter meta algoritmoma gozdov (Breiman, 1996) in naključnih gozdov (Breiman, 2001) (preglednica 2).

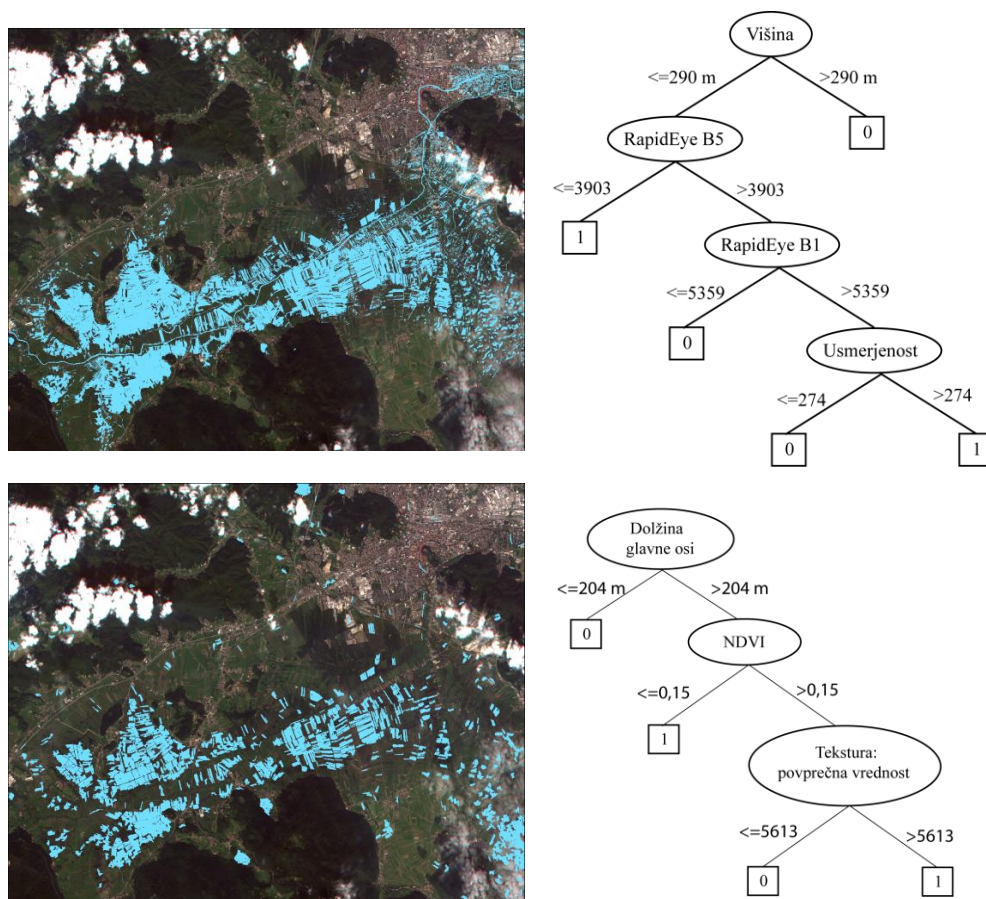
5.2.3 Rezultati

Predstavljeni so rezultati uspešnosti klasifikacije glede na obliko učnega vzorca in uporabljen algoritem strojnega učenja. Primeri posameznih klasifikacij so tudi grafično predstavljeni v obliki kart.

Rezultat učenja na vzorcu točk oz. segmentov

Uspešnost klasifikacije z uporabo algoritma J48 je bila 89 % v primeru uporabe učnega vzorca s točkami in 85 % v primeru učnega vzorca s segmenti. Uspešnost klasifikacije ob uporabi učnega vzorca točk je torej 4 % višja kot v primeru uporabe učnega vzorca segmentov. Kot se vidi na sliki 20, je prepoznavanje poplavljenih površin v primeru uporabe učnih točk boljše zlasti med polji Ljubljanskega barja. Težavo povzročajo urbana območja, kjer prihaja do zamenjav pozidanih površin s poplavljenimi. Na drugi strani pri klasifikaciji s segmenti ne prihaja do zamenjav pozidanih površin s poplavljenimi, podcenjena pa je zaznava poplavljenih površin med polji na Ljubljanskem barju, kot poplavljenega pa so napačno prepoznana območja rahle oblačnosti v JV delu posnetka. V modelu za klasifikacijo, pridobljenim z učenjem na učni množici točk, imajo največji vpliv na klasifikacijo spektralna atributa modrega in bližnje IR kanala ter atributa višine in usmerjenosti. Klasifikacijo s segmenti usmerjajo indeks NDVI, dolžina glavne osi in tekstura (odločitveni drevesi na sliki 20). To so najznačilnejši atributi za določitev poplavljenih površin na tipično strukturiranem območju Ljubljanskega barja, na katerem prevladujejo dolge in ozke njive z vmesnimi pasovi ali gručami gozdov. Poplavljenega tla pod gozdom in visokimi kulturnimi rastlinami niso zaznana, zaznane so le poplavljenе vmesne površine, ki so zaradi lege med polji prav tako značilnih pravilnih geometrijskih oblik, od ostalih površin pa se razlikujejo po teksturi. Natančnosti klasifikacij s segmenti, ki sta bili izvedeni v programu ENVI, sta bili še nekoliko nižji, 83 % v primeru klasifikacije z algoritmom podpornih vektorjev oz. 82 % v primeru klasifikacije z uporabo najbližjega sosedu.

Rezultati so pokazali, da učenje s točkami na tipično strukturiranem površju Ljubljanskega barja omogoča natančnejšo klasifikacijo poplavljenih površin kot učenje na primeru segmentov. Preizkus ostalih algoritmov za učenje v programu WEKA in CLUS je zato potekal na učni množici točk. Natančnosti vseh klasifikacij, tako tistih, ki so bile opravljene na osnovi učenja s točkami, kot tiste, ki so potekale z učenjem na segmentih, so prikazane v nadaljevanju.



Slika 20: Klasifikacija poplavljenih površin z uporabo algoritma J48 na učnem vzorcu točk (zgoraj) in na učnem vzorcu segmentov (spodaj). Na desni strani sta prikazana modela v obliki odločitvenega drevesa.

Figure 20: Classification of flooded areas using algorithm J48 with training sample of points (above) and with training sample of segments (below). Decision tree of each classification is shown on the right.

Ocena natančnosti klasifikacije z uporabo različnih algoritmov strojnega učenja

Uspešnost učenja s posameznim algoritmom je bila določena z uporabo 10-kratnega navzkrižnega preverjanja na učnem vzorcu, sestavljenem iz 201 točke oz. segmenta, ocena natančnosti klasifikacije pa je potekala na osnovi 100 testnih točk, katerih atribut poplavljenosti je bil določen s pomočjo vizualne interpretacije letalskih posnetkov DOF IR, ki so nastali dan po koncu obilnih padavin 20. 9. in fotografske dokumentacije z dne 20. 9. in 23. 9. (Foto: Matija Zorn). Natančnosti klasifikacij z uporabo različnih algoritmov strojnega učenja so predstavljene v preglednici 13 za primere treh klasifikacij s segmenti in preglednici 14 za primere petih klasifikacij s točkami.

Oceno natančnosti klasifikacije poplavljenih površin je zaradi specifičnosti območja, na katerem se pojavljajo polja z visokimi rastlinskimi vrstami in območja z drevesi, težko oceniti. Vzemimo za primer koruzno polje ne daleč stran od struge Ljubljance. Omenjeno polje je na satelitskem posnetku vidno kot nepoplavljen del ozemlja, saj kuruza sega nad gladino vode. Z vidika spektralnih atributov satelitskega posnetka je območje upravičeno klasificirano kot nepoplavljeno. Težava pa se pojavi ob upoštevanju atributov oddaljenosti od vodotokov in atributov digitalnega modela višin. Majhna

oddaljenost od reke, nizka nadmorska višina in ravno dno Ljubljanskega barja ustrezajo lastnostim poplavljenih površin. V takšnih primerih se torej pojavi konflikt, ali omenjeno polje uvrstiti v razred poplavljenih ali nepoplavljenih površin. Naša raziskava je usmerjena predvsem v zmožnost prepoznavanja poplavljenih površin s satelitskih posnetkov, zato imajo prednost atributi satelitskih posnetkov. Polja z visoko vegetacijo in poplavljenim dnom tako označimo kot nepoplavljena. Uvrstitev omenjenega polja v razred poplavljenih površin štejemo za napako, čeprav bi bila z vidika atributov, ki ne pripadajo spektralnim lastnostim satelitskega posnetka, tudi takšna klasifikacija lahko pravilna.

Preglednica 13: Uspešnost določitve poplav z uporabo različnih algoritmov strojnega učenja, ki je potekalo na učnem vzorcu, sestavljenem iz segmentov.

Table 13: Accuracy of determination of flooded areas using different machine learning methods with the training set of segments.

Metoda:	Delež pravilno klasificiranih primerov	
	Učenje	Test
J48	95 %	85 %
SVM	/	83 %
NN	/	82 %

Preglednica 14: Uspešnost določitve poplav z uporabo različnih algoritmov strojnega učenja, ki je potekalo na učnem vzorcu, sestavljenem iz točk.

Table 14: Accuracy of determination of flooded areas using different machine learning methods with the training set of points.

Metoda:	Delež pravilno klasificiranih primerov	
	Učenje	Test
J48	95 %	89 %
JRip	96 %	88 %
Gozd J48 (10 trees)	95 %	83 %
Naključni gozd C45 (5 trees)	98 %	80 %
Naključni gozd C45 (10 trees)	96 %	82 %

Najvišjo natančnost (89 %) dosega klasifikacija z uporabo algoritma J48 in učenjem na učni množici točk. Natančnost ostalih klasifikacij z uporabo vzorca točk je malenkost nižja pri uporabi algoritma pravil JRip (88 %) in slabša pri uporabi gozdov oz. naključnih gozdov (80 % do 83 %). Vse klasifikacije, ki so potekale z učenjem na učnem vzorcu segmentov, so dosegale zelo podobne natančnosti, ki se gibljejo med 82 % in 85 %. Za najuspešnejšo klasifikacijo, pridobljeno z algoritmom J48, je prikazana tudi matrika napak (preglednica 15), ki omogoča določitev izdelovalčeve in uporabnikove natančnosti. Ti znašata 84 % (izdelovalec) oz. 93 % (uporabnik) za razred poplavljenih površin ter 94 % (izdelovalec) oz. 85 % (uporabnik) za razred nepoplavljenih površin. Nižja

izdelovalčeva natančnost poplavljenih površin je posledica neprepoznanih poplavljenih površin, ki so bile napačno klasificirane kot nepoplavljen del površja. Razlog je predvsem v prostorski ločljivosti posnetka RapidEye (6,5 m), ki je za zaznavanje ozkih poplavljenih območij med koruznimi polji prenizka.

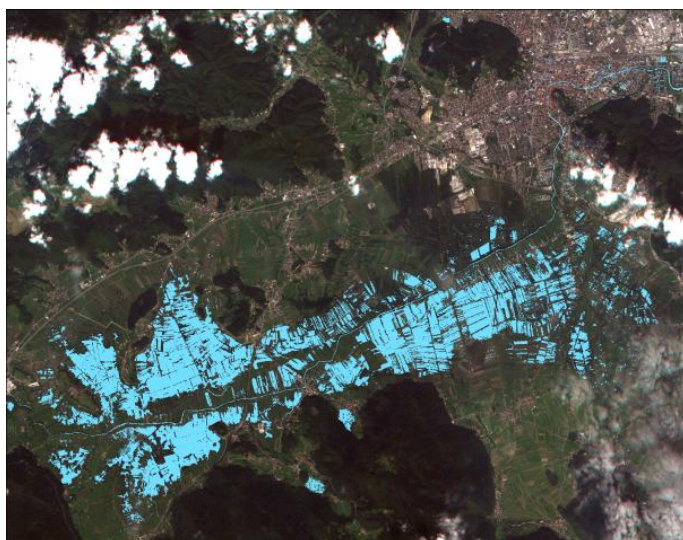
Preglednica 15: Matrika napak za oceno natančnosti klasifikacije na območju Ljubljanskega barja.

Table 15: Confusion matrix to define classification accuracy in the study area of Ljubljana moor.

Razred	Poplavljeno	Nepoplavljen	Klasificirane točke	Natančnost (%)	
				Izdelovalec	Uporabnik
Poplavljeno	42	3	45	84	93
Nepoplavljen	8	47	55	94	85
Referenčne točke	50	50	100		

Kartiranje poplavljenih površin

V nadaljevanju so prikazane karte poplavljenih površin na območju Ljubljanskega barja (slika 21). Prikazane karte se razvrščene v vrstnem redu od najslabšega do najboljšega rezultata. Ob vsaki posamezni karti so pojasnjene tudi glavne posebnosti, ki vplivajo na uspešnost zaznave poplav. Učni vzorci vseh prikazanih klasifikacij so bili sestavljeni iz točk.

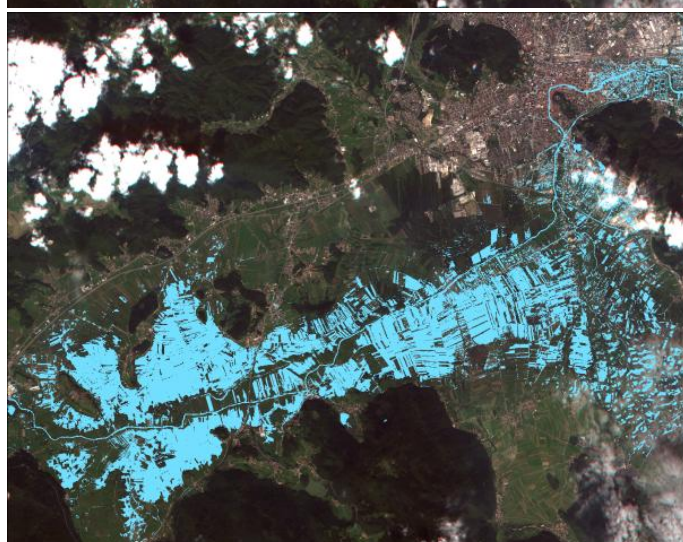
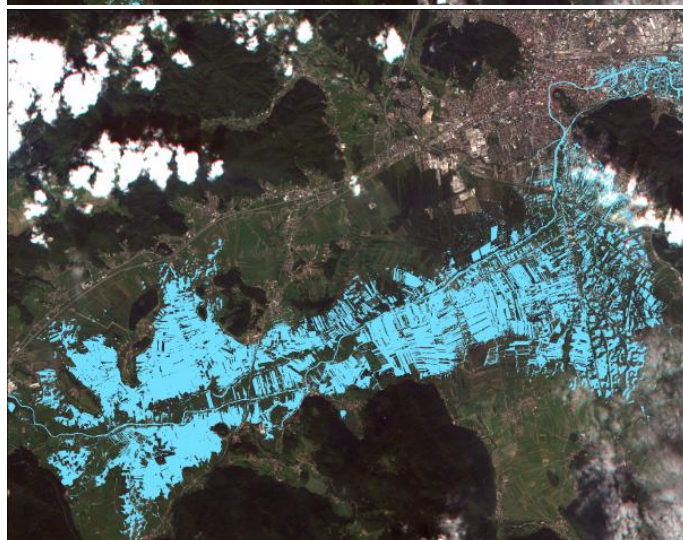
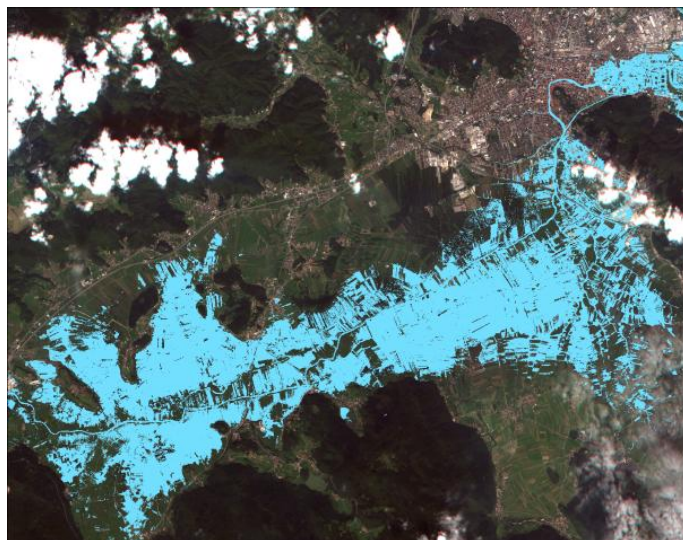


Naključni gozd

- Obseg poplavljenega območja je podcenjen na kmetijskih površinah.
- Zamenjav urbanih površin s poplavljenimi je zelo malo.

Model: Množica z desetimi drevesi je prikazana v prilogi (Priloga E).

Se nadaljuje...



...nadaljevanje.

Gozd

- Obseg poplavljenega območja je bolje določen na kmetijskih površinah, kjer je kot poplavljen

prepoznan tudi precejšen del poplavljenih površin pod visoko vegetacijo.

- Veliko pozidanih površin na območju Ljubljane je napačno prepoznano kot poplavljen.

Model: Množica z desetimi drevesi je prikazana v prilogi (Priloga D).

JRip

- Zelo dobro so prepoznane poplavljenе površine izven urbanega območja.

- Težavo še vedno predstavljajo urbana območja.

Model v obliki odločitvenih pravil:

1. (DMV ≤ 289 m) in (B5-bližnja IR ≤ 3711) $\Rightarrow$ voda = 1
2. (DMV ≤ 290 m) in (B1-modra ≥ 5203) in (usmerjenost ≥ 248) $\Rightarrow$ voda = 1
3. sicer $\Rightarrow$ voda = 0

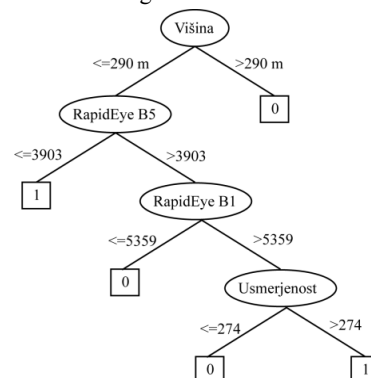
J48

Uspešnost določitve poplavljenih površin podobna kot v primeru algoritma JRip:

- dobro so prepoznane poplavljenе površine izven urbanega območja,

- težavo predstavljajo urbana območja.

Model v obliki odločitvenega drevesa:



Slika 21: Kartiranje poplavljenih površin z uporabo različnih algoritmov strojnega učenja na učnem vzorcu sestavljenem iz točk.

Figure 21: Maps of flooded areas produced by different machine learning algorithms and training sample of points.

Med klasifikacijami z različnimi vrstami algoritmov prihaja do razlik v uspešnosti klasifikacije zlasti na kmetijskih in urbanih površinah. Presenetljiva je slabša natančnost klasifikacij z uporabo množice odločitvenih dreves. Z algoritmom naključnih gozdov je podcenjena količina poplavljenega površja na kmetijskih površinah, uspešno pa so prepoznana urbana območja kot nepoplavljena. Ravno obratno so z algoritmom gozdov uspešneje določena poplavljena ozemlja na kmetijskih površinah in slabše na urbanih območjih. Na drugi strani je klasifikacija, ki potekala z uporabo enega samega modela, pridobljenim z algoritmom JRip oz. z algoritmom J48, uspešnejša od uporabe množic. V obeh primerih uporabe enega samega modela je rezultat zelo podoben. Klasifikacija izven urbanih površin je dobra, med pozidanimi območji pa prihaja podobno kot v primeru metode gozdov do zamenjav s poplavljenimi površinami.

Kartiranje poplav na časovni vrsti radarskih posnetkov

Doktorska disertacija je osredotočena na uporabo optičnih večspektralnih satelitskih posnetkov, ki omogočajo zaznavanje odbitega valovanja v različnih kanalih EMV. Informacije o odbitem valovanju različnih kanalov med sabo kombiniramo, zaradi česar lahko bolje ločimo in prepoznavamo predmete na zemeljskem površju. Na drugi strani senzorji radarskih sistemov delujejo precej bolj samostojno. Zaznavajo različne valovne dolžine odbitega mikrovalovanja, ki omogočajo zbiranje podatkov o kopnem, vodi, ledu ali atmosferi. Radarski posnetki so bili uporabljeni izključno za izgradnjo časovne vrste posnetkov in spremljanje postopnega umikanja vode iz poplavljenih površin na Ljubljanskem barju. Ugodna časovna vrsta, sestavljena iz enega posnetka Envisat in treh posnetkov Radarsat (preglednica 16), nam je omogočila izdelavo karte spreminjanja obsega poplavljenih površin v obdobju 14 dni po poplavah (slika 22). Zadostnega števila optičnih posnetkov za izgradnjo časovne vrste posnetkov za obdobje po poplavah nismo uspeli pridobiti. Oviro pri uporabi večjega števila optičnih posnetkov znotraj kratkega časovnega obdobja predstavljajo predvsem oblaki, skozi katere zemeljsko površje na optičnih posnetkih ni vidno. Za pravilno interpretacijo radarskih posnetkov je potrebno dobro razumevanje načina delovanja radarskih sistemov, kar je opisano v poglavju 2.1.1. Vrednosti piksla na radarskih posnetkih predstavljajo intenziteto povratnega signala, ki je odvisna od razgibanosti/hrapavosti in dielektričnosti površja, na katerem je prišlo do odboja. Določitev poplavljenih površin iz radarskih posnetkov je potekala po metodi določevanja pragu (thresholding) (Singh, 1989), ki predstavlja iskanje mejne vrednosti odbitega EMV med dvema različnima razredoma. V našem primeru prag predstavlja mejno vrednost med odbojem radarskega valovanja na poplavljenih in nepoplavljenih površinah. Metoda določevanja pragu je uporabljena v različnih aplikacijah za prepoznavanje objektov na zemeljskem površju. Metoda pa ne omogoča samodejne določitve pragu z zadostno natančnostjo, ampak je za določitev prave mejne vrednosti potrebno veliko ponovitev s poizkušanjem različnih vrednosti (Liew in sod., 1998; Dierking in Skriver, 2002). Delo je zelo zamudno zlasti v primeru uporabe večjega števila podatkovnih slojev/posnetkov, ko je potrebno mejno vrednost določiti za vsak atribut posebej (Itami in sod., 2004). V tem primeru je priporočljivejša

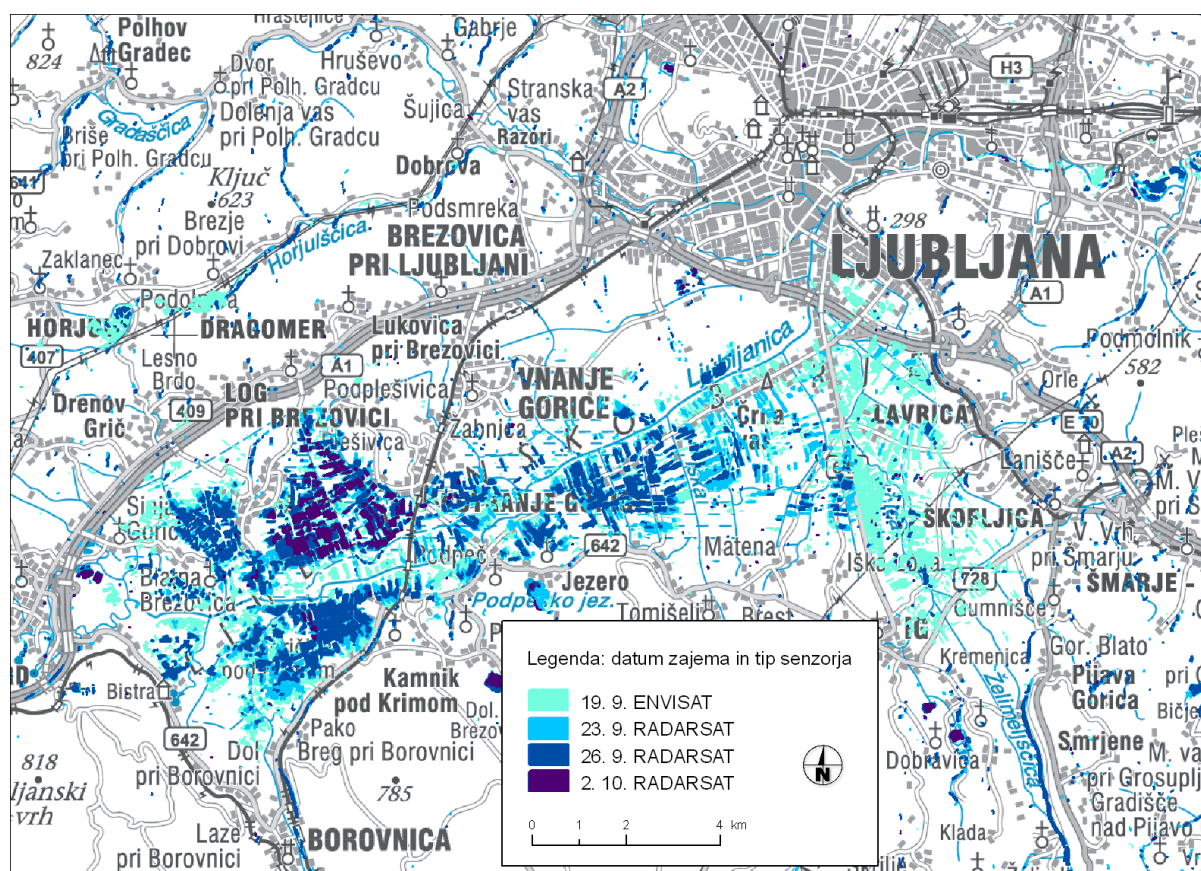
uporaba tehnike strojnega učenja, ki omogočajo določitev značilnih vrednosti vseh atributov istočasno. S tem pa je čas, ki ga porabimo za izgradnjo klasifikacijskega modela, krajši, poskrbeti pa moramo za pripravo ustreznega učnega vzorca, saj je od kakovosti učnega vzorca v veliki meri odvisna tudi uspešnost strojnega učenja.

Preglednica 16: Časovna vrsta uporabljenih radarskih satelitskih posnetkov.

Table 16: List of time series of the radar satellite images to observe flooded areas in time.

Satelit (lastnik)	Datum zajema	Prostorska ločljivost (m)
Envisat (ESA)	19. 9. 2010	12,5
Radarsat-2 (CSA)	23. 9. 2010	12,5
Radarsat-2 (CSA)	26. 9. 2010	12,5
Radarsat-2 (CSA)	2. 10. 2010	12,5

Najprej so bila določena poplavljenega območja na posameznem radarskem posnetku z metodo pragu. Iz posameznih kart poplavljenih površin je bila nato izdelana časovna vrsta posnetkov, ki omogoča dober vpogled v obseg in dinamiko poplav. Slika 22 prikazuje prekrivanje štirih radarskih satelitskih posnetkov, iz katere se lepo vidi postopno umikanje vode s poplavljenega ozemlja.



Slika 22: Prikaz umikanja vod s prekrivanjem štirih radarskih satelitskih posnetkov: enega posnetka Envisat in treh posnetkov Radarsat. Prikaz na državni topografski karti DTK 50 (GURS, 2005).

Figure 22: Overlaying of four radar images indicates how the floodwater was receding from affected areas. Presentation on digital topographic map DTK 50 (GURS, 2005).

5.2.4 Razprava

Analiza možnosti prepoznavanja poplavljenih površin na območju Ljubljanskega barja je prinesla zelo podobne ugotovitve, kot smo jih spoznali že na območju Železnikov, pa tudi nekaj bistvenih razlik. Raziskava je pokazala, da so atributi, ki usmerjajo klasifikacijo s posameznimi modeli, hkrati najpomembnejši tudi pri klasifikacijah z uporabo množic odločitvenih dreves. Med njimi na območju Ljubljanskega barja izstopata le atribut naklona in četrtega kanala posnetka RapidEye (rob rdeče), ki pomembno vplivata na klasifikacijo z metodo naključnih gozdov, v preostalih primerih pa ne sodelujeta (preglednica 17). Glede na primer Železnikov se je spremenila vrsta najpomembnejših atributov. Odločitveno drevo na sliki 20 (zgoraj) predstavlja najnatančnejši model klasifikacije na območju Ljubljanskega barja. V klasifikacijski model so vključeni atributi višin, bližnji IR in modri kanal posnetka RapidEye in usmerjenost. Na začetku klasifikacije nadmorska višina najprej izloči nepoplavljena višja območja v okolici Ljubljanskega barja, ki jih poplave ne dosežejo. Sledi upoštevanje odbojnosti v IR delu spektra, ki je visoka na vegetaciji in nizka na vodnih površinah, kar nam je omogočilo jasno razlikovanje gozdnih in kmetijskih površin (npr. koruzna polja) od poplavljenih območij. V drugem delu drevesa nastopata še modri kanal in atribut usmerjenost, ki pomagata pri natančnejšem določevanju poplavljenih površin predvsem na urbanih območjih, vendar je njihovo ločevanje zaradi prevladujočega ravninskega površja in le nekoliko višje odbojnosti v modrem delu spektra na pozidanih površinah glede na poplavljen tla še vedno zelo zahtevno. Bližnji IR kanal se je kot pomemben atribut izkazal že pri klasifikaciji na območju Železnikov, medtem ko senzorji satelitskega sistema SPOT podatka o odbitem valovanju v modrem delu vidne svetlobe ne omogočajo. Razlog za drugačno sestavo atributov, ki opisujejo oblikovanost površja, na obeh študijskih območjih, je drugačna oblikovanost površja. Na Ljubljanskem barju prevladuje obsežno ravninsko območje, na katerem se vode, ki prestopijo robove rečnih strug, razlijejo po širši okolici. Oddaljenost od vodotokov zato nima več tako pomembnega vpliva na določitev poplavljenih površin kot na območju Železnikov. Prav tako nima večjega vpliva naklon, ki se je izkazal kot zelo pomemben na območju Železnikov. Kot pomembna atributa na območju Ljubljanskega barja pa sta se izkazala še atributa višin in usmerjenosti, ki nudita informacijo o oblikovanosti površja in s tem vplivata na določitev poplavljenih območij.

Preglednica 17: Prikaz atributov, ki sodelujejo pri klasifikacijah poplavljenih površin na območju Ljubljanskega barja z učenjem na učnem vzorcu točk in različnimi algoritmi. Vrednosti v preglednici predstavljajo število modelov, v katerih se pojavi posamezen atribut. Večje število pojavljanj pomeni večji vpliv atributa na klasifikacijo.

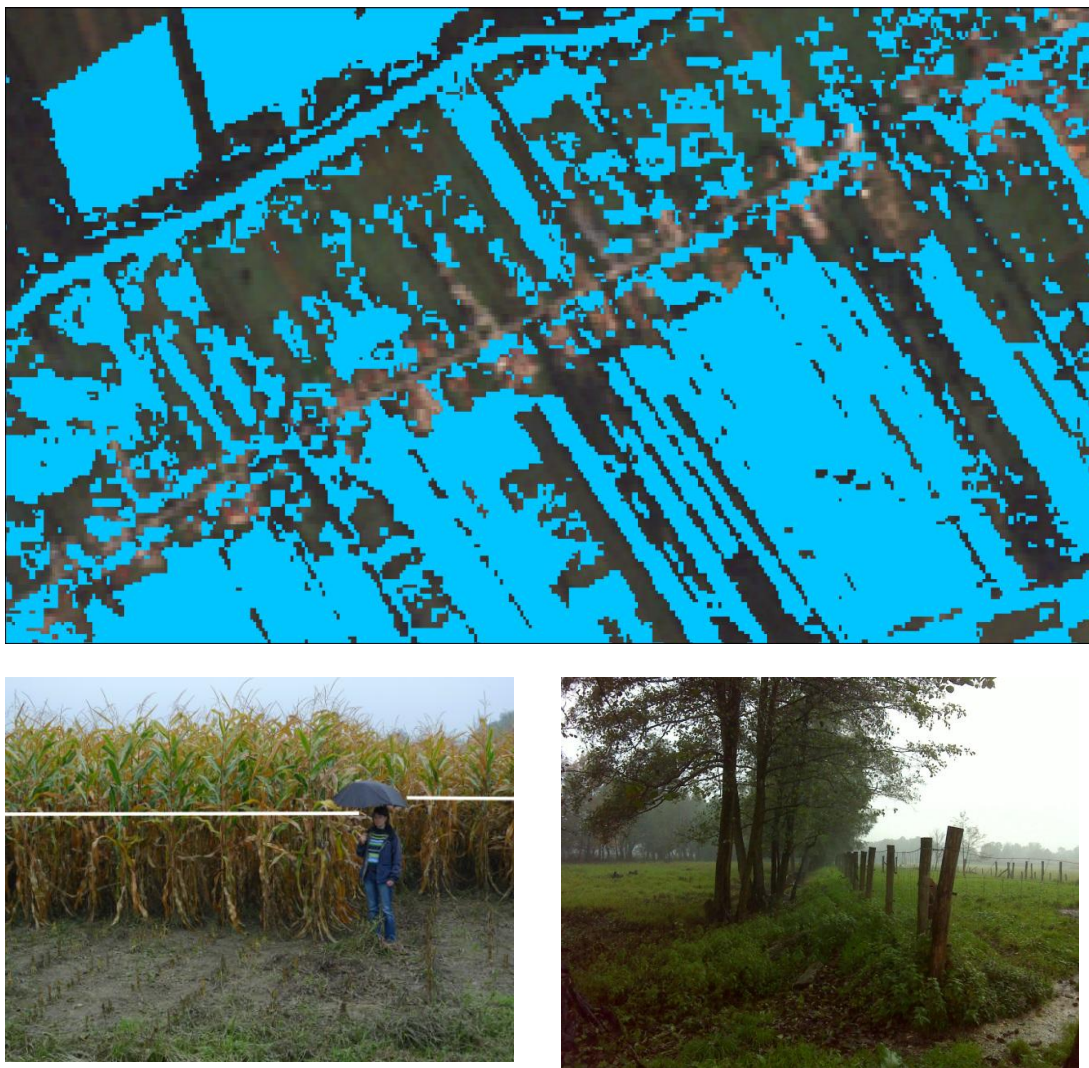
Table 17: Attributes included into the model for classification of flooded areas in the area of Ljubljana Moor. The learning process was performed using sample of points and different algorithms. Values in the table present the number of models where the attributes were taking part. Larger values of the attributes show higher impact on the classification process.

Atribut	Gozd*	Naključni gozd*	J48	JRip
B1 – moder	3	3	1	1
B2 – zelen	0	1	0	0
B3 – rdeč	1	2	0	0
B4 – rob rdeče	0	6	0	0
B5 – bližnja IR	9	7	1	1
NDVI	1	3	0	0
Višina	10	8	1	1
Naklon	0	7	0	0
Usmerjenost	2	5	1	1
Ukrivljenost	1	2	0	0
Oddaljenost od rek	0	1	0	0

*Metodi gozdov in naključnih gozdov sta potekali z izgradnjo 10 dreves.

Najnatančnejša določitev poplavljenih površin je bila pridobljena z uporabo algoritma za učenje z odločitvenimi drevesi J48 in učnim vzorcem točk. Segmenti sicer prispevajo dodatne informacije o objektih na zemeljskem površju (npr. geometrijo in teksturo), ki jih točkam ni mogoče določiti in ki olajšajo prepoznavanje poplavljenih površin, vendar so bili v našem primeru segmenti preveliki, zaradi česar je bila natančnost klasifikacije z njimi nižja. Za učenje smo poskušali uporabiti čim manjše segmente, vendar zaradi gladke teksture in značilne geometrijske oblike poplavljenih površin, ki je posledica lege med koruznimi polji, manjše velikosti segmentov nismo uspeli doseči. Posledično je količina poplavljenega površja, določenega s segmenti, nekoliko podcenjena, saj so prepoznane le večje sklenjene poplavljenе površine. Težava je še večja zaradi prostorske ločljivosti posnetkov RapidEye (6,5 m), ki je zelo podobna širini pasov poplavljenih površin med koruznimi polji, kar lahko privede do priključitve manjših in ožjih poplavljenih območij k sosednjim segmentom nepoplavljenih površin.

Pri zaznavanju poplavljenih površin na Ljubljanske barju poseben primer predstavljajo visoke kulturne rastline, predvsem koruzna polja, ki ne omogočajo zaznave poplavljenih tal pod njimi. Podoben učinek imajo tudi skupine dreves in območja gozdov. Ta območja imajo na satelitskih posnetkih vse značilnosti nepoplavljenega ozemlja, zato jih je tako potrebno upoštevati tudi v postopku učenja. Posledica je značilna nesklenjena razporeditev poplavljenih površin, ki jih prekinjajo vmesna območja kopnega dela površja (slika 23).



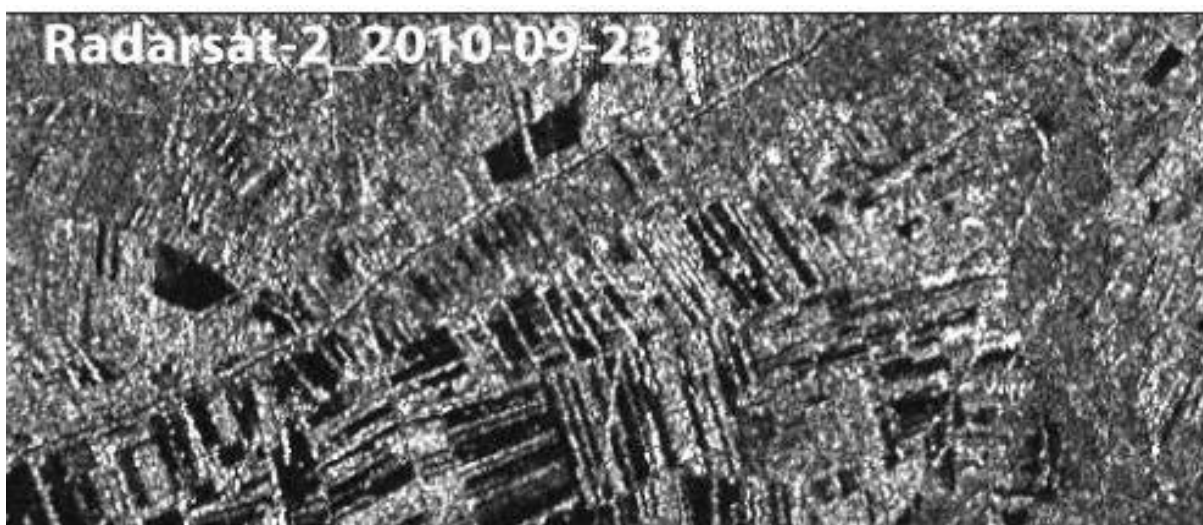
Slika 23: Prepoznavanje poplavljenih površin med objekti, ki segajo nad gladino vode. Poplavljene površine so označene z modro barvo (slika zgoraj). Nepoplavljen del predstavljajo višji objekti, ki segajo nad gladino vode, npr. koruznega polja in drevesni nasad (fotografiji spodaj).

Figure 23: Flooded areas between high objects presented with blue colour (above). Non-flooded areas are higher objects extending above water surface, e.g. corn fields, trees (photos below).

Težave prepoznavanja poplavljenih površin so se pojavljale tudi na urbanih območjih, na katerih je prav tako prihajalo do zamenjav pozidanih površin s poplavljenimi. Nepravilnosti v zaznavanju poplavljenih površin so povezane s prostorsko ločljivostjo satelitskega posnetka RapidEye (6,5 m) in digitalnega modela višin (DMV 12,5), ki sta prenizki za zaznavo posameznih hiš na gosto pozidanih območjih. Težavo povzročajo tudi zelo podobne spektralne lastnosti poplavljenih površin in pozidanih območij. Nevarnost zamenjave je še toliko večja zaradi različnih pojavnih oblik, ki jih lahko zavzamejo vodne površine in so posledica radiometrične nehomogenosti poplavljenih površin (slika 19), vzvalovanosti vodne gladine in različne globine. Te vplive smo poskušali zmanjšati s postopki filtriranja (glajenja), kljub temu pa vseh vplivov ni bilo mogoče odpraviti.

Težav senc na radarskih posnetkih, ki so preprečevale uporabo radarskih posnetkov na območju Železnikov, na ravninskem območju Ljubljanskega barja ni bilo, kar je omogočilo njihovo uporabo za

izgradnjo časovne vrste in spremljanje sprememb obsega poplavljenih površin. Podobno kot na optičnih posnetkih RapidEye so težavo na radarskih posnetkih povzročali nad vodno gladino segajoči objekti. Zahtevnost določitve poplavljenega površja je še večja zaradi tipične podolgovate in ozke oblike koruznih polj. Njihova širina je v nekaterih primerih podobna velikosti piksla na radarskem posnetku (12,5 m), kar zmanjšuje natančnost zaznanih poplavljenih površin. Na radarskih posnetkih je vidna tudi zrnatost, ki se kaže v pikčasti strukturi posnetka in daje občutek posejanosti posnetka s »poprom in soljo« (Oštir, 2006), vendar ta na zaznavo poplavljenih površin nima bistvenega vpliva (slika 24).

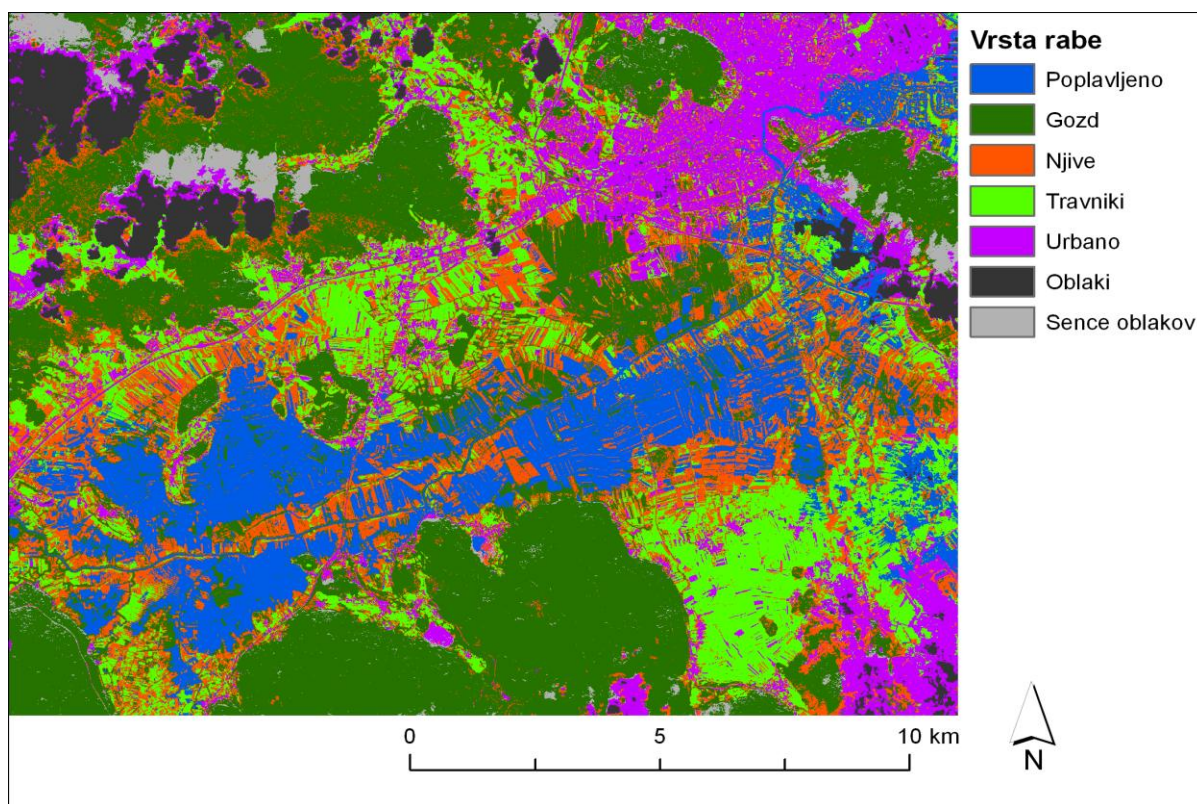


Slika 24: Poplavljene površine vidne na radarskem posnetku Radarsat. Poplavljeno območje ni strnjeno zaradi vmesnih polj in dreves. Vidna je tudi zrnatost – posejanost posnetka s črno-belimi drobnimi strukturami (»posejanost s poprom in soljo«).

Figure 24: Flooded areas shown on Radarsat image. Flooded areas are not homogeneous because of fields and trees. The image is affected by speckle noise.

Izdelana je bila tudi karta, ki prikazuje vrsto rabe na obravnavanem ozemlju (slika 25). Za učenje je bila uporabljena ista množica točk kot pri vseh ostalih klasifikacijah, le učnim točkam je bil dodan atribut rabe tal. Klasifikacija vključuje sedem razredov: voda, gozd, njive, travniki, urbane površine, oblaki in sence oblakov. V razred vod se poleg učnih točk, ki ležijo na območju vodotokov (npr. reka Ljubljana), uvrščajo tudi vse točke na poplavljenih površinah. Točke, ki ležijo na nepoplavljenih območjih, so razvrščene v preostalih šest razredov. Model za klasifikacijo s sedmimi razredi je bil zgrajen na osnovi učenja z metodo gozdov z 10 drevesi, ki pri dani količini podatkov zagotavlja maksimalno uspešnost, z nadaljnjim povečevanjem števila odločitvenih dreves pa se natančnost klasifikacije prične zmanjševati. S pomočjo karte rabe je interpretacija kart poplavljenih površin precej lažja, saj je s karte razvidno, da osrednji del Ljubljanskega barja, ki je na kartah poplavljenosti zaznan kot nepoplavljen, prekrivajo njive, polja, deloma tudi gozd. Iz tega lahko sklepamo, da je tudi na tem površju večina tal poplavljenih, ampak zaznavo vode preprečuje nad vodno gladino segajoča vegetacija. S primerjavo kart poplavljenih površin in karte rabe lahko prepoznamo tudi prekomerno

zaznane poplavljene površine, ki so se na območju Ljubljanskega barja pojavljale na urbanih območjih.



Slika 25: Karta rabe na območju Ljubljanskega barja. Namen karte je lažja interpretacija kart poplavljenih površin in prepoznavanje težav, ki preprečujejo bolj strnjeno zaznavo poplavljenih površin.

Figure 25: Map of land use in the area of Ljubljana Moor to make interpretation of flood maps easier.

6 MOŽNOSTI UPORABE REZULTATOV IN NADALJNJE DELO

Rezultat prepoznavanja poplavljenih površin po predstavljenem postopku je karta poplavljenega ozemlja, katere praktična vrednost je predvsem v lažjem in hitrejšem postopku ocenjevanja ob poplavah nastale škode. Natančnost klasifikacije je bila v našem primeru zadostna za oceno škode zlasti na kmetijskih površinah, medtem ko je bila natančnost zaznanih poplavljenih površin med stavbami na urbanih območjih premajhna. Višja natančnost je bila dosežena z uporabo učnega vzorca točk, vendar tudi uporaba segmentov, kljub nekoliko nižji uspešnosti klasifikacije, zadostuje za praktično uporabo. V praksi bi bilo zelo koristno, v kolikor bi lahko za klasifikacijo poplavljenih površin uporabili razdelitev na segmente, kot jo določa katera od uradnih evidenc podatkov, npr. evidenca dejanske rabe kmetijskih in gozdnih zemljišč. Evidenco je izdelalo takratno Ministrstvo za kmetijstvo, gozdarstvo in prehrano (MKGP), njene podatke pa danes redno vzdržujejo pod okriljem Ministrstva za kmetijstvo in okolje. Evidenca je nastala zaradi potreb Registra kmetijskih gospodarstev (RKG) in nadzora nad dodeljevanjem finančnih spodbud (subvencij) (Mivšek in sod., 2012). Klasifikacija z uporabo segmentov, kot jih določa ta evidenca, bi omogočila še hitrejšo oceno nastale škode. Nadaljnja primerjava poplavljenih segmentov z grafičnimi enotami rabe zemljišč kmetijskih gospodarstev (GERK) (MKGP, 2009) pa bi omogočila neposredno oceno nastale škode za vsako posamezno kmetijsko gospodarstvo. Uporaba segmentov, kot jo določa evidenca dejanske rabe kmetijskih in gozdnih zemljišč, je bila sicer preizkušena, vendar se je izkazalo, da njenih segmentov ni mogoče uporabiti v postopku učenja, saj so med njimi številni segmenti le deloma poplavljeni. Kmetijsko zemljišče ni vedno poplavljeno v celoti, ampak je lahko del zemljišča poplavljenega in del nepoplavljenega. Tako je potrebno za potrebe klasifikacije poplavljenih površin izvesti samostojno segmentacijo, ocena nastale škode za posamezno kmetijsko gospodarstvo pa lahko poteka s prekrivanjem karte poplavljenih površin in GERK-ov.

Za dosego višje natančnosti, ki bi omogočila oceno škode tudi na pozidanih površinah, bi potrebovali podatke (predvsem satelitske posnetke) višje prostorske ločljivosti. Možnosti za izboljšanje postopka prepoznavanja poplavljenih površin obstajajo tudi v uporabi še večjih količin različnih vrst podatkov in preizkusu dodatnih tehnik strojnega učenja. Nadaljnje delo bo tako vključevalo podatke, ki tokrat niso bili uporabljeni. Največ se pričakuje od uporabe natančnejšega digitalnega modela višin in novih satelitskih posnetkov z boljšimi spektralnimi in prostorskimi lastnostmi. Digitalni model višin, izdelan iz lidarskih podatkov, zagotavlja višjo prostorsko ločljivost in boljšo višinsko natančnost, kar bi omogočilo modeliranje pretakanja vod in s tem izboljšanje natančnosti določitve poplavljenih površin (Hunter in sod, 2008). Poleg tega lidarski podatki nudijo tudi informacijo o ostalih objektih na zemeljskem površju (npr. vegetaciji in grajenih objektih), v okolici katerih je določevanje poplavljenosti najzahtevnejše. S satelitskimi posnetki z višjo spektralno in prostorsko ločljivostjo (večjim številom kanalov) bi pridobili večje število spektralnih lastnosti opazovanega dela

zemeljskega površja, kar bi izboljšalo razlikovanje med opazovanimi objekti. Z višjo prostorsko ločljivostjo bi zaznali tudi manjše objekte, vendar pa se v teh primerih lahko na satelitskih posnetkih pojavi šum, ki oteži prepoznavanje poplavljenih površin. Za Ljubljansko barje so trenutno že na voljo lidarski podatki, ki so nastali v okviru projekta laserskega skeniranja celotne Slovenije (GURS, 2011) in ki v času poteka raziskave še niso bili na razpolago. Izmed vseh trenutno delujočih optičnih satelitskih sistemov pa so za opazovanje poplav med najbolj primernimi posnetki Pleiades. Pleiades predstavlja konstelacijo dve optičnih satelitov za opazovanje zemeljskega površja z zelo visoko ločljivostjo, ki sta pričela delovati v letih 2011 oz. 2012 (Baillarin in sod., 2009). Kombinacija dveh satelitov prinaša številne prednosti, kot sta možnost sistematičnega dnevnega zajema podatkov za izbrano območje zemeljskega površja in v primeru hkratne usmerjenosti obeh satelitov na isto območje zajem precej bolj obsežnih površin. Dva satelita povečata tudi možnost pridobivanja satelitskih posnetkov brez prisotne oblačnosti. Pleiades predstavlja nadaljevanje satelita SPOT. Ima zelo ugodno sestavo spektralnih kanalov za opazovanje v modrem, zelenem, rdečem, bližnje IR ter pankromatskem kanalu in v primerjavi s satelitskih sistemom SPOT veliko boljšo prostorsko ločljivost (50 cm). Za leto 2014 se načrtuje pričetek obratovanja še enega optičnega satelita Sentinel-2, ki bo v primerjavi s Pleiades omogočal snemanje z nižjo prostorsko ločljivostjo (10–60 m), slabšo časovno ločljivostjo (2–3 dni) in boljšo spektralno ločljivostjo (13 spektralni kanalov v vidnem, bližnjem IR in kratkovalovnem IR delu spektra) (Martimort in sod., 2012). Uporaba novih satelitov predstavlja nove možnosti za izboljšanje prepoznavanja poplavljenih površin na detajlnih območjih. Tako pričakujemo, da bo z novimi satelitskimi posnetki v prihodnje mogoče izboljšati tudi prepoznavanje poplav na urbanih območjih in na območjih z visoko vegetacijo, ki so za oceno škode bistvenega pomena.

Nekoliko nižja stopnja uporabnosti kart poplavljenih površin se predvideva za potrebe reševanja in nudenja prve pomoči, saj na območju Slovenije prevladujejo poplave manjših razsežnosti, zaradi česar so v uporabi predvsem letalski posnetki. Karte poplavljenih površin, izdelane s satelitskih posnetkov, pa so vsekakor uporabne za potrebe reševanja v primerih naravnih nesreč večjih razsežnosti, kar je tudi eden od ciljev uporabe satelitskih posnetkov Pleiades in Sentinel-2 (Boissin in sod., 2012; Martimort in sod., 2007). Določevanje poplavljenih površin omogoča tudi zagotavljanje večje varnosti pred poplavami in načrtovanje protipoplavnih ukrepov (Sivapalan in sod., 2003). Analiza posamezne poplave še ne prinaša nujno zadovoljivih ugotovitev, pač pa je potrebno večletno analiziranje vseh poplavnih dogodkov. Šele rezultati več analiz omogočajo njihovo primerjavo in sklepanje trdnejših ugotovitev. Npr. hidravlični parametri, ki so bili izračunani na osnovi karte poplavljenih površin na območju Železnikov (Lamovec in Mikoš, 2011), se lahko uporabijo kot vir podatkov za umerjanje hidroloških modelov. Za takšno hidravlično analizo pa potrebujemo digitalni model višin visoke natančnosti, ki omogoča natančno določitev višine vode poplavljenega območja. Analize poplav so potrebne za razumevanje hidrološkega odziva določenega območja na intenziteto, porazdelitev in

količino padavin, kar je nujno za uspešno napovedovanje poplav in pravočasno opozarjanje prebivalstva pred njimi.

7 ZAKLJUČEK

V okviru doktorske disertacije smo na osnovi eksperimentalnih primerov izdelali postopek za hitro in natančno izdelavo karte poplavljenih površin, ki je namenjen ocenjevanju nastale škode, hidrološkim analizam in učinkovitejši organizaciji nudenja pomoči prizadetemu prebivalstvu. Za doseg navedenih ciljev potrebujemo kakovostne podatke in učinkovit postopek njihove obdelave, kar je v raziskavi doseženo s kombinacijo večspektralnih optičnih satelitskih posnetkov in strojnega učenja. Dosedanji postopki prepoznavanja poplavljenih površin so potekali na osnovi klasičnih postopkov klasifikacije večspektralnih satelitskih posnetkov, to je nadzorovane in nenadzorovane klasifikacije ali vizualne interpretacije. V naši raziskavi pa smo postopek učenja in postopek klasifikacije razdelili na dva povsem samostojna dela, kar predstavlja povsem nov pristop k prepoznavanju poplavljenih površin. Za razliko od ustaljenih postopkov klasifikacij, ki potekajo z izbiro učnega vzorca in razvrščanjem preostalih primerov v razrede po izbrani metodi klasifikacije, postopek klasifikacije z uporabo strojnega učenja omogoča natančno analizo vplivov ločljivosti DMV, gostote učnega vzorca, oblike učnih primerov in različnih algoritmov na uspešnost učenja, kar zagotavlja višjo natančnost klasifikacije. Skladno z namenom izboljšanja postopka prepoznavanja poplavljenih površin so bile oblikovane tri raziskovalne hipoteze, ki predstavljajo vodilno nit skozi celotno doktorsko disertacijo:

1. Z uporabo strojnega učenja je mogoče izboljšati hitrost in doseči višjo natančnost določitve poplavljenih površin, kot jo zagotavljajo klasični postopki klasifikacije z večspektralnimi satelitskimi posnetki.
2. Relief in pokrovnost tal imata velik vpliv na uspešnost zaznave poplavljenega površja. Vpliv pokrovnosti je mogoče z uporabo strojnega učenja zmanjšati.
3. S časovno vrsto satelitskih posnetkov je mogoče opazovati dinamiko umikanja vode iz poplavljenih površin.

V raziskavi smo preizkusili različne vplive na uspešnost učenja in natančnost klasifikacije. Rezultat eksperimentov je določitev postopka strojnega učenja, ki omogoča najnatančnejšo določitev poplavljenih površin glede na ločljivost DMV, gostoto učnega vzorca, obliko učnih primerov in vrsto algoritma strojnega učenja. Postopek je bil preizkušen na dveh tipih poplav: na poplavah s hudourniškim značajem na območju Železnikov in na poplavah s kraškim značajem na območju Ljubljanskega barja. Obravnava obeh tipov poplav, ki se pojavljata na območjih z različnimi značilnostmi površja, je pokazala razlike v klasifikaciji, zaradi katerih ju ne moremo obravnavati enako. Rezultati so pokazali, da najboljšo kombinacijo za pridobitev modela za klasifikacijo poplavljenih površin tako na razgibanih območjih s hudourniškiimi poplavami kot na območjih kraških poplav predstavlja kombinacija učenja z vzorcem točk in algoritmom J48. Razlika pa se pojavlja v atributih, ki usmerjajo klasifikacijo na obeh študijskih območjih. Pri klasifikaciji, ki dosega najvišjo natančnost, so na območju hudourniških poplav v Železnikih najpomembnejši zeleni kanal, bližnji

infrardeči kanal, oddaljenost od vodotokov in naklon, v primeru kraškega tipa poplav na Ljubljanskem barju pa so to nadmorska višina, bližnji infrardeči kanal, modri kanal in usmerjenost. Spektralni kanali, uporabljeni v obeh primerih so podobni. Tako je v obeh primerih v klasifikacijo vključen bližnji IR kanal, ki ima na poplavljenih površinah zaradi absorpcije bližnje IR svetlobe na vodi zelo nizke vrednosti. Odboj modre in zelene svetlobe pa se na vodi poveča, zaradi česar so vrednosti odboja v teh dveh kanalih višje. Razlike pa se pojavijo v atributih, ki opisujejo oblikovanost površja in oddaljenost od vodotokov. Tako na območju Železnikov pomembno vlogo igrata atributa naklon in oddaljenost od vodotokov. Na območju Ljubljanskega barja omenjene atributa nimata vpliva na klasifikacijo, saj je mreža vodotokov tu zelo razvejana, mreža vodnih kanalov pa se razteza po celotnem obravnavanem območju. V klasifikaciji pa sodelujeta atributa višina in umerjenost, ki sta na ravninskem območju Ljubljanskega barja precej bolj konstanta kot v primeru Železnikov. Naklon na Ljubljanskem barju nima večjega vpliva zaradi ravninskega območja in tudi zaradi uporabe digitalnega modela višin slabše natančnosti.

Predhodno določeni najbolj primerni atributi omogočajo hitrejšo izgradnjo modela za klasifikacijo poplavljenih površin ob podobnem dogodku, saj omogočajo pripravo podatkovnih slojev, ki jih potrebujemo za določitev poplavljenih površin. Ob poplavnem dogodku pa učenim primerom že pripravljenega učnega vzorca ponovno določimo le atribut poplavljenosti, ki določa položaj točk na poplavljenih oz. nepoplavljenih tleh. Pričakovati je, da je zgradba modelov za prepoznavanje poplavljenih površin na drugih območjih, podvrženih hudourniškem oz. kraškemu tipu poplav, podobna, vendar je za učinkovitost postopka na celotnem območju Slovenije potrebna priprava zadostnega števila modelov, ki bodo omogočili pravilno določitev poplavljenih površin na poljubnem območju v realnem času.

Razviti postopek klasifikacije poplavljenih površin je sicer le eden izmed korakov v celotnem postopku izdelave karte poplavljenih površin, ki se prične z zajemom satelitskega posnetka z različnimi sistemi in pripravo ostalih podatkov, nadaljuje s predobdelavo (geometrijski in radiometrični popravki), izračunom dodatnih podatkovnih slojev (NDVI) in konča s kartiranjem poplavljenih površin. V okviru centra odličnosti Vesolje-SI razvita procesna veriga za hitro kartiranje, katere del je tudi v doktorski disertaciji predstavljen postopek prepoznavanja poplavljenih površin, bo omogočala izdelavo kart poplavljenih površin v nekaj urah po sprožitvi snemanja prizadetega območja z izbranim satelitskim sistemom, kar je bistveno hitreje od ostalih sistemov za hitro kartiranje, ki omogočajo izdelavo kart v približno enem dnevu.

S kombiniranjem več različnih vrst podatkov v postopku strojnega učenja smo izboljšali natančnost določitve poplavljenih površin glede na natančnosti dosežene z uporabo klasičnih postopkov klasifikacije. Klasifikacija poplavljenih površin na Ljubljanskem barju je z uporabo nadzorovane klasifikacije večspektralnega satelitskega posnetka RapidEye je dosegla za 7 % nižjo natančnost, kot je bila natančnost v primeru klasifikacije po modelu, pridobljenim s strojnim učenjem z algoritmom

J48. Glede na hitrost postopka določevanja poplavljenih površin, ki je del procesne verige za avtomatsko izdelavo kart v skoraj realnem času, in glede na dosežene natančnosti klasifikacij, lahko prvo hipotezo v celoti potrdimo.

Dosežene natančnosti klasifikacij poplavljenih površin na Ljubljanskem barju so v splošnem nekoliko nižje od natančnosti poplavljenih površin v Železnikih, kar je posledica zahtevnejših pogojev za prepoznavanje poplav na Ljubljanskem barju. Na obeh študijskih območjih je bila najuspešnejša klasifikacija pridobljena z uporabo algoritma J48 na učnem vzorcu točk, njena natančnost je 95 % na območju Železnikov oz. 89 % na območju Ljubljanskega barja. Nižja natančnost klasifikacije na območju Ljubljanskega barja je predvsem posledica večjega deleža površin, ki ga zavzemajo nad vodno gladino segajoči objekti, predvsem drevesa in visoke poljščine na kmetijskih površinah. Težavo pa povzročajo tudi zamenjave pozidanih površin s poplavljenimi. Atributi, pridobljeni iz DMV, ne omogočajo ločitve visokih objektov od poplavljenih površin, saj vsebujejo le informacijo o oblikovanosti zemeljskega površja. Ločevanje med površinami, pokritimi z vegetacijo in poplavljenimi površinami, pa omogočajo spektralne vrednosti satelitskega posnetka RapidEye. Vrednosti modrega kanala so višje na poplavljenih površinah, na katerih prihaja do večjega odboja modre svetlobe. Vrednosti bližnjega IR kanala pa so višje na območjih vegetacije, medtem ko se večina bližnje IR svetlobe na poplavljenih površinah absorbira. Na območju Železnikov težavo povzročajo podobne spektralne vrednosti satelitskega posnetka SPOT na poplavljenih, pozidanih in kmetijskih površinah, vendar je območij, na katerih prihaja do zamenjav, manj kot na Ljubljanskem barju. Z natančno izbiro učnih primerov, ki so bili razporejeni na za prepoznavanje poplav zelo zahtevnih območjih (npr. krošnje dreves, strehe hiš), smo izboljšali natančnost klasifikacije tudi na teh površinah.

Boljše rezultate klasifikacije smo pričakovali z uporabo segmentov, ki pa v primerjavi z uporabo učnega vzorca s točkami ne zagotavljajo izboljšanja natančnosti klasifikacije niti na Ljubljanskem barju, niti na območju Železnikov. Neuspešnost klasifikacije s segmenti na območju Železnikov je posledica podobnih lastnosti, ki jih imajo poplavljeni segmenti in segmenti na nepoplavljenih urbanih ter kmetijskih površinah. Na Ljubljanskem barju težave pri ločevanju segmentov, ki pripadajo poplavljenim površinam, od segmentov na vegetaciji, povzroča tipična strukturiranost površja, na katerem prevladujejo dolge in ozke njive, ki vplivajo na podobno obliko tudi poplavljenih površin.

Na območju Železnikov se je kot zelo pomemben podatek za pravilno določitev poplavljenih površin izkazal relief. Uporaba DMV 0,5, ki je bil izdelan s tehniko laserskega skeniranja, je na območju Železnikov vplivala na zvišanje natančnosti klasifikacije za 3 % v primerjavi z uporabo DMV 12,5. Na osnovi predstavljenih rezultatov lahko drugo hipotezo potrdimo le deloma. Potrdil se je pomemben vpliv digitalnega modela višin na natančnost klasifikacije, saj v primeru uporabe DMV višje natančnosti dosežemo boljše rezultate klasifikacije. V nasprotju s pričakovanji, pa z uporabo segmentov nismo uspeli zmanjšati vpliva pokrovnosti na uspešnost klasifikacije, ampak se je z njihovo

uporabo natančnost klasifikacije znižala. Kljub temu pa se je na območju Železnikov potrdilo, da je s pravilno izbiro učnega vzorca, ki mora vključevati tudi primere iz vode segajočih objektov in kombiniranjem več različnih vrst podatkov, tudi podatkov o reliefu, mogoče izboljšati natančnost določitve poplavljenih površin tudi na zahtevnejših območjih.

Doktorska disertacija je bila osredotočena na uporabo optičnih večspektralnih satelitskih posnetkov, vendar večjega števila optičnih posnetkov v krajšem časovnem intervalu za izgradnjo časovne vrste nismo imeli na razpolago, saj so bili številni posnetki zaradi prekritosti z oblaki neuporabni. Izgradnja časovne vrste je bila zato mogoča samo iz radarskih posnetkov, ki omogočajo opazovanje zemeljskega površja praktično neodvisno od vremena, torej tudi v primeru oblačnosti. Štirje radarski satelitski posnetki Envisat in Radarsat na območju Ljubljanskega barja, nastali znotraj časovnega razmika 14 dni, so omogočili spremljanje dinamike umikanje vod, kar potrjuje tretjo hipotezo. Karta, nastala s prekrivanjem poplavljenih površin, ki so bile določene iz uporabljenih radarskih posnetkov, nazorno prikazuje, iz katerih območij je voda najprej odtekla in na katerih območjih se je zadržala najdlje. Z natančnejšimi višinskimi podatki o oblikovanosti površja, pa bi lahko ocenili tudi količino vode in hitrost njenega odtekanja.

Kljub temu da je strojno učenje na nekaterih področjih že v rutinski rabi, je na področju prepoznavanja poplavljenih površin še dokaj neizkoriščeno. Prepoznavanje poplavljenih površin iz optičnih satelitskih posnetkov s tehnikami strojnega učenja je bilo do danes preizkušeno z uporabo posameznih algoritmov. Doktorska disertacija pa predstavlja celosten preizkus in primerjavo različnih algoritmov strojnega učenja. Na dveh tipih poplav so bili opisani vsi koraki od priprave podatkov do njihove klasifikacije. Določeni so bili vpliv prostorske ločljivosti DMV na uspešnost učenja, pomembnost izbire reprezentativnega učnega vzorca in posebnosti, ki se pojavljajo pri hudourniškem oz. kraškem tipu poplav. Rezultati so pokazali najnatančnejšo določitev poplavljenih površin z uporabo algoritma J48 na učnem vzorcu točk, ki mora biti dovolj velik, za zajem vseh raznolikost obravnavanega območja.

Postopek določitve poplavljenih površin z uporabo daljinskega zaznavanja in strojnega učenja je bil razvit in preizkušen na dveh poplavnih dogodkih, ki predstavljata tudi različni tip poplav (hudourniški in kraški). Nekateri značilnosti poplav so skupne vsem poplavnim dogodkom, druge so značilne le za posamezen tip poplave, hkrati pa je vsak poplavni dogodek enkraten pojav, ki se razlikuje od vseh ostalih in je v povsem enaki obliki neponovljiv. Rezultati doktorske disertacije predstavljajo primerno osnovo za opazovanje podobnih dogodkov tako v Sloveniji kot drugod. Vsaka klasifikacija pa potrebuje kljub znani okvirni sestavi klasifikacijskega modela za določen tip poplav na izbranem ozemlju določene prilagoditve, ki pa so s postopki strojnega učenja ob upoštevanju rezultatov doktorske raziskave hitro in učinkovito opravljene.

8 POVZETEK

Poplave so v zadnjih letih najpogostejše naravne nesreče, za posledicami katerih trpi največ svetovnega prebivalstva, prinašajo veliko gmotno škodo na kmetijskih površinah, infrastrukturi, stanovanjskih in gospodarskih poslopih ter v najhujših primerih zahtevajo tudi človeška življenja. Namen doktorske disertacije je bil pokazati in ovrednotiti postopek, ki bo zagotavljal hitro in natančno prepoznavanje poplavljenih površin. Za ustrezno natančnost potrebujemo dovolj kakovostne podatke in za primerno hitrost ustrezen postopek njihove obdelave. Končni rezultat je karta poplavljenih površin, ki jo lahko pri svojem delu uporabijo službe, ki se ukvarjajo z reševanjem na prizadetih območjih, ocenjevanjem nastale škode in določevanjem poplavno najbolj izpostavljenih območij.

Doktorska disertacija združuje dve področji, katerih velik napredek v razvoju sta omogočila razvoj računalniške in digitalne tehnologije, to sta področji daljinskega zaznavanja in strojnega učenja. Z napredkom tehnologije so se povečale baze prostorskih podatkov, katerih pomemben del predstavljajo podatki, pridobljeni s pomočjo sistemov daljinskega zaznavanja. Na voljo je vedno več podatkov o oblikovanosti površja, rabi tal, infrastrukturi idr., ki se razlikujejo v prostorskih, spektralnih in časovnih lastnostih. To pa je bil tudi motiv za preizkus uporabe strojnega učenja za potrebe prepoznavanja poplavljenih površin.

Strojno učenje je omogočilo iz uporabljenih podatkov pridobiti najpomembnejše informacije, ki so potrebne za določitev poplavljenih površin. Največji del porabljenega časa v postopku strojnega učenja je potrebno nameniti naboru različnih vrst podatkov in sestavi ustreznega učnega vzorca. Tako so v doktorski disertaciji predstavljeni rezultati strojnega učenja z učnimi vzorci, ki se razlikujejo po gostoti oz. številu učnih primerov, njihovi obliki in atributih vključenih v postopek učenja. Na osnovi ocenjene natančnosti klasifikacije z uporabo različnih učnih vzorcev je bil določen najuspešnejši učni vzorec, ki je bil nato uporabljen za preizkus različnih algoritmov strojnega učenja.

Učni vzorci in algoritmi strojnega učenja so bili preizkušeni na dveh študijskih območjih, območju Železnikov in Ljubljanskega barja, ki se med sabo zelo razlikujeta po geomorfoloških značilnostih. Prvo območje predstavlja hudourniško dolino ujeto med vrhove Škofjeloškega hribovja in Bohinjskih gora, drugo študijsko območje je ravninsko območje z značilnostmi kraškega tipa poplav. Atributi, ki so sodelovali pri določitvi poplavljenih površin na območju Železnikov, so bili pankromatski kanal satelitskega posnetka SPOT, večspektralni kanali satelitskega posnetka SPOT ločljivosti 10 m in 2,5 m, indeksi NDVI, NBI, NDBI, višina, naklon, usmerjenost in oddaljenost od vodotokov. Tudi na Ljubljanskem barju so se ponovno kot zelo pomembne izkazale spektralne lastnosti satelitskega posnetka, le da je bil namesto posnetka SPOT tokrat uporabljen satelitski posnetek RapidEye. Ostali uporabljeni atributi so bili še indeks NDVI, višina, naklon, usmerjenost, ukrivljenost in oddaljenost od vodotokov.

Na območju Železnikov sta bila uporabljena dva učna vzorca z različnim številom učnih primerov. Uporabljena sta bila učna vzorca z 255 učnimi točkami, razporejenimi po celotnem obravnavanem območju, in s 145 učnimi točkami, razporejenimi na manjšem izseku, s čimer je ta učni vzorec dosegal višjo gostoto učnih točk. Višja gostota zagotavlja natančnejši opis obravnavanega območja, kar se je izkazalo kot zelo pomembno, saj se je natančnost klasifikacije v primeru uporabe učnega vzorca s 145 točkami zvišala za 5 % in je znašala 86%.

Na obeh območjih pa sta bila preizkušena učna vzorca dveh različnih oblik, sestavljena iz točk oz. segmentov. Obakrat se je kot učinkovitejši izkazal učni vzorec s točkami, ki je z uporabo modela, pridobljenega z algoritmom odločitvenih dreves J48, zagotavljal 4 % višjo natančnost klasifikacije kot v primeru uporabe segmentov. Na območju Železnikov je tako natančnost klasifikacije z učnim vzorcem točk znašala 95 % in na območju Ljubljanskega barja 89 %. Učni vzorec točk je bil nato uporabljen tudi za preizkus algoritmov strojnega učenja z odločitvenimi pravili JRip ter metodama za kombiniranje klasifikatorjev z gozdovi in naključnimi gozdovi, vendar natančnost klasifikacije z ostalimi algoritmi ni presegla natančnosti odločitvenih dreves J48. Na območju Železnikov nekoliko slabši rezultat zagotavljajo klasifikacije, izvedene po metodi za kombiniranje klasifikatorjev (v obeh primerih 92 %), medtem ko je uspešnost klasifikacije z odločitvenimi pravili precej nižja (85 %). Nekoliko drugačna je bila učinkovitost posameznih algoritmov na območju Ljubljanskega barja, kjer so odločitvena pravila JRip zagotavljala le en odstotek nižjo natančnost klasifikacije (88 %) od najvišje natančnosti klasifikacije, dosežene z uporabo algoritma odločitvenih pravil J48, medtem ko sta bili klasifikaciji z metodama gozdov in naključnih gozdov precej nižji (83 % in 82 %).

9 SUMMARY

Floods are the most frequent natural disaster in the last years that affect the largest number of people. They cause an immense damage on agricultural areas, infrastructure, apartments and industrial buildings and in the most critical cases they cause also human lives. The aim of the doctoral dissertation is to show and evaluate the process for faster and more accurate detection of flooded areas. High quality input data and suitable process for their analysis is needed for best results. Final product of the processing is a map of flooded areas which could be used by emergency services to support people in affected areas, to make damage assessment and determine the most exposed areas to the floods.

Doctoral dissertation combines remote sensing and machine learning, two areas where development accelerated by great progress in computer and digital technology. Machine learning was applied in economy, medicine and other fields where big amount of data has to be managed effectively. In recent years the amount of data in spatial database is increasing rapidly. The most important part of these is data produced by remote sensing. More and more data about Earth surface, land use, infrastructure etc. is available in different spatial, spectral and temporal resolutions. That was the motivation of the thesis and to use machine learning methods for flooded areas determination.

With machine learning it is possible to extract the most important information that is necessary to define flooded areas. The most time in machine learning process was spent to collect different kind of data and to find the suitable training sample. In the thesis results of machine learning with different training samples are presented. The main differences are the density and the number of training examples, their shape and attributes included in the learning process. The most successful training sample was defined by accuracy assessment of different samples and was used for testing different machine learning techniques further.

Training samples and machine learning algorithms were tested in two study areas: Železniki and Ljubljana Moor. Both study areas are very different according to geomorphological features. The first one is presented by mountain valley surrounded by Škofja Loka Mountains and Bohinj Mountains. The second is presented by flat swamp area with characteristics of karst type of the floods. The classification of flooded area in Železniki was directed by attributes of panchromatic and multispectral SPOT satellite images of resolution 10 and 2,5 m, indexes NDVI, NBI, NDBI, height, slope, aspect and distance from water. In the area of Ljubljana Moor spectral attributes of SPOT satellite images used in the case of Železniki were replaced by attributes of RapidEye satellite image. Index NDVI, height, slope, aspect, curvature and distance from water were considered again.

Two training samples were tested using different number of training examples in the area of Železniki. The training sample with 255 training points arranged in the whole treated area and 145 training points arranged in the smaller part of the area were tested in the classification. The density of training sample with 145 points arranged in the smaller part of the area was higher and describes treated area more precisely. It makes possible to classify flooded areas in higher accuracy. The accuracy of classification in the area of Železniki was 81 % using 255 training points and increased for 5 % on 86 % using 145 training points.

In both areas Železniki and Ljubljana Moor training samples of points and segments were also tested using decision tree algorithm J48. In both cases the training sample of points was proven as more sufficient. The accuracy of classification was 4 % higher using training sample of points compared with training sample of segments. The classification accuracy was 95 % in the area of Železniki and 89 % in the area of Ljubljana Moor using the training sample of points. Considering the accuracy of classifications with different training samples points were chosen for learning with other machine learning algorithms: decision rules algorithm JRip and two ensemble methods bagging and random forest. Their classification accuracy did not exceed the accuracy of classification achieved by decision tree learning algorithm J48. In the area of Železniki lower classification accuracy was acquired by ensemble methods bagging and random forest (92 % in both examples) and the lowest classification accuracy was produced by decision rule algorithm JRip (85 %). In the area of Ljubljana Moor the classification accuracy of decision rule algorithm JRip was just 1 % (88 %) lower than the highest classification accuracy was produced by decision tree algorithm J48. Classification accuracy of two ensemble methods bagging and random forest were quite lower (83 % and 82 %).

10 VIRI

- Alpaydin, E. 2010. Introduction to Machine Learning (Adaptive Computation and Machine Learning). Second Edition. Massachusetts Institute of Technology. Cambridge: The Mit Press.
- ARSO. 2010. Poročilo o izjemno obilnih padavinah od 16. –19. sep. 2010. Ljubljana: Državna meteorološka služba, 22. 9. 2010.
- Ashraf, M.A., Maah, M.J., Yusoff, I. 2011. Introduction to Remote Sensing of Biomass. V: Atazadeh, I. (ur.). Biomass and Remote Sensing of Biomass.
<http://www.intechopen.com/books/biomass-and-remote-sensing-of-biomass/introduction-to-remote-sensing-of-biomass> (Pridobljeno 10.3.2013.)
- Ayala, A.I. 2002. Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. *Geomorphology* 47: 107–124.
- Ayanu, Y.Z., Conrad, C., Nauss, T., Wegmann, M., Koellner, T. 2012. Quantifying and Mapping Ecosystem Services Supplies and Demands: A Review of Remote Sensing Applications. *Environmental Science & Technology* 46: 8529–8541.
- Bach, H., Dierschke, U., Appel, F., Fellah, H., de Fraipont, P. 2004. Application of satellite data for flood monitoring. V: Proceedings GGRS, Göttingen/Germany. Frankfurt a. M., J. D. Sauerländer Verlag.
- Bach, H., Appel, F., Fellah, K., de Fraipont, P. 2005. Application of flood monitoring from satellite for insurances. IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2005.
- Baillarin, S., Lebégue, L., Kubik, P. 2009. Pleiades-HR system qualification: A focus on ground processing and image products performances, a few months before launch. IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2009: I-76–I-79.
- Berk, R.A. 2005. Data mining within a regression framework. V: Maimon, O. (ur.), Rokach, L. (ur.). Data mining and knowledge discovery handbook. New York, Springer Science & Business: 231–255.
- Blaschke, T., Strobl, J. 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS-Zeitschrift für Geoinformations systeme* 14 (6), 12–17.
- Blaschke, T., Lang, S., Hay, G.J. 2008. Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications. Berlin Heidelberg, Springer-Verlag: 836 str.
- Blaschke, T. 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65, 2–16.
- Boissin, M.B., Gleyzes, A., Tinel, C. 2012. The pléiades system and data distribution. IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2012: 7098–7101.
- Boutsinas, B. 2002. Accessing data mining rules through experts systems. *International Journal of Information Technology and Decision Making* 1, 735: 657–672.
- Breiman, L. 1996. Bagging Predictors. *Machine Learning* 24, 2: 123–140.

- Breiman, L. 2001. Random Forests. *Machine Learning* 45, 1: 5–32.
- Burgan, R.E., Hartford, R.A. 1996. Live vegetation moisture calculated from NDVI and used in fire danger rating. 13th Conference on Fire and Forest Meteorology, 27 October 1996, Fairfield WA, IAWF: 225–231.
- Chen, P., Liew, S.C., Lim, H. 1999. Flood detection using multitemporal Radarsat and ERS SAR data. V: Proc. of the 20th ACRS, 22–25 November, 1999, HongKong.
- Chen, S.T., Yu, P.S. 2007. Pruning of support vector networks on flood forecasting. *Journal of Hydrology* 347, 1–2: 67–78.
- Chen, J., Li, M., Liu, Y., Shen, C., Hu, W. 2010. Extract residential areas automatically by New Built-up Index. *Geoinformatics*, 18th International Conference, 18–20 June 2010, Beijing.
- Cohen, W.W. 1995. Fast effective rule induction. V: Proc. of the 12th International Conference on Machine Learning, Tahoe City, CA. San Francisco, Morgan Kaufmann: 115–123.
- Degiorgis, M., Gnecco, G., Gorni, S., Roth, G., Sanguineti, M., Taramasso, A.C. 2012. Classifiers for the detection of flood-prone areas using remote sensed elevation data. *Journal of Hydrology* 470–471: 302–315.
- De Moel, H., van Alphen, J., Aerts J. C. J. H. 2009. Flood maps in Europe – methods, availability and use. *Nat. Hazards Earth Syst. Sci.* 9: 289–301.
- Dierking, W., Skriver, H. 2002. Change detection for thematic mapping by means of airborne multitemporal polarimetric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing* 40, 3: 618–636.
- Dietterich T.G. 2000. Ensemble Methods in Machine Learning. *Lecture Notes in Computer Science* 1857: 1–15.
- Džeroski, S., Kocev, D. 2012. Pomoč pri uporabi ansambelskih metod v programu CLUS. Osebna komunikacija (oktober 2011).
- Estivill-Castro, V., Yang, J. 2000. A Fast and robust general purpose clustering algorithm. *Pacific Rim International Conference on Artificial Intelligence*: 208–218.
- Fabbricatore, C., Boley, H., Karduck, A.P. 2012. Machine learning for resource management in smart environments. 6th IEEE International Conference on Digital Ecosystems Technologies (DEST): 1–6.
- Faruolo, M., Coviello, I., Lacava, T., Pergola, N., Tramutoli, V. 2009. Real time monitoring of flooded areas by a multi-temporal analysis of optical satellite data. *IEEE International Geoscience and Remote Sensing Symposium*, , IGARSS 2009: IV-192–IV-195.
- Frappart, F., Minh, K.D., L'Hermitte, J., Cazenave, A., Ramillien, G., Toan, T.L., Campbell, N.M. 2006. Water volume change in the lower Mekong from satellite altimetry and imagery data. *Geophysical Journal International* 167, 2: 570–584.
- Gentleman, R., Carey, V.J. 2008. Unsupervised Machine Learning. *Bioconductor Case Studies*: 137–157.

- Glenn, N.F., Streutker, D.R., Chadwick, D.J., Glenn, D.T., Dorsch S.J. 2006. Analysis of lidar-derived topographic information for characterizing and differentiating landslide morphology and activity. *Geomorphology* 73, 1–2: 131–148.
- Goldman, S., Zhou, Y. 2000. Enhancing supervised learning with unlabeled data. *Proceedings of the 17th International Conference on Machine Learning*, Morgan Kaufman, San Francisco, CA: 327–334.
- Grabowski, S., Jóźwik, A., Chen, C.H. 2003. Nearest neighbor decision rule for pixel classification in remote sensing. *Frontiers Of Remote Sensing Information Processing*: 315–327.
- GURS. 2011. Lasersko skeniranje in aerofotografiranje 2011. Predstavitev projekta na tiskovni konferenci dne 3. 6. 2011. Ministrstvo za okolje in prostor.
http://www.gu.gov.si/fileadmin/gu.gov.si/pageuploads/novice/Teksti_novic/LIDAR_opis.pdf
(Pridobljeno 16. 11. 2012)
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H. 2009. The WEKA Data Mining Software: An Update; *SIGKDD Explorations* 11, 1.
- Han, S.W., Kim, J.Y. 2008. Rough set-based decision tree using a core attribute. *International Journal of Information Technology and Decision Making* 7, 2: 275–290.
- Heckerman, D. 1999. A Tutorial on Learning with Bayesian Networks. V: Jordan, M. (ur.). *Learning in Graphical Models*. MIT Press, Cambridge, MA.
- Henderson, F.M., Lewis, A.J. 1998. *Principles and Applications of Imaging Radar*, 3rd ed. New York: American Society for Photogrammetry and Remote Sensing.
- Herbrich, R. 2001. *Learning Kernel Classifiers: Theory and Algorithms*. MIT Press: 375 str.
- Hsu, C.W., Chang, C.C., Lin, C.J. 2007. A practical guide to support vector classification. National Taiwan University.
<http://ntu.csie.org/~cjlin/papers/guide/guide.pdf> (Pridobljeno 8. 12. 2011)
- Hunter, N.M., Bates, P.D., Neelz, S., Pender, G., Villanueva, I., Wright, N.G., Liang, D., Falconer, R.A., Lin, B., Waller, S., Crossley, A.J., Mason, D.C. 2008. Benchmarking 2D hydraulic models for urban flooding. *Proceedings of the ICE - Water Management* 161, 1: 13–30.
- Irimescu, A., Craciunescu, V., Stancalie, G., Nertanet, A. 2010. Remote Sensing and GIS Techniques for Flood Monitoring and Damage Assessment. Study Case in Romania. *BALWOIS 2010*. Ohrid, Republic of Macedonia.
- Itami, F., Watanabe, E., Nishihara, A. 2004. A change detection method for image sequences based on physical models. *The 47th Midwest Symposium on Circuits and Systems, MWSCAS '04*: I-293–I-296.
- Jemec, M., Mikoš, M. 2008. Pobočni masni premiki na satelitskih posnetkih SPOT: Primer območja Železnikov po vodni ujmi septembra 2007. *Geologija* 51, 2: 235–243.
- Kocev, D., Vens, C., Struyf, J., Dzeroski, S. 2013. Tree ensembles for predicting structured outputs. *Pattern Recognition* 46, 3: 817–833.

- Kokalj, Ž., Marsetič, A., Pehani, P., Oštir, K. 2008. Observation of torrential rains devastation in Slovenia. V: Proceedings of the 1st International Conference on Remote Sensing Techniques in Disaster Management and Emergency Response in the Mediterranean Region. Zagreb, Croatian Academy of Sciences and Arts, HAZU: 181–190.
- Komac, B., Ciglič, R., Erhartič, B., Gašperič, P., Kozina, J., Orožen Adamič, M., Pavšek, M., Pipan, P., Volk, M., Zorn, M. 2010. Risk education and natural hazards. CapHaz-Net WP6 Report, Anton-Melik Geographical Institute of the Scientific Research Centre of the Slovenian Academy of Sciences and Arts: Ljubljana.
http://caphaz-net.org/outcomes-results/CapHaz-Net_WP6_Risk-Education2.pdf (Pridobljeno 15. 3. 2012).
- Kotsiantis, S.B. 2007. Supervised Machine Learnig: A Review of Classification Techniques. V: Maglogiannis, I.G. (ur.), Karpouzis, K. (ur.), Wallace, M. (ur.), Soldatos, J. (ur.). Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies. IOS Press: 3–24.
- Krzysztofowicz, R. 2001. Integrator of uncertainties for probabilistic river stage forecasting: precipitation-dependent model. *Journal of Hydrology*, 249: 69–85.
- Krzysztofowicz, R. 2002. Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology*, 268: 16–40.
- Laliberte, A.S., Goforth, M.A, Steele, C.M., Rango, A. 2011. Multispectral Remote Sensing from Unmanned Aircraft: Image Processing Workflows and Applications for Rangeland Environments. *Remote Sens.* 3, 11: 2529–2551.
- Lamovec, P., Oštir, K. 2010. Uporaba strojnega učenja za določitev poplavljenih območij - primer poplav v Selški dolini leta 2007. *Geodetski vestnik* 54, 4: 673–687.
- Lamovec, P., Mikoš, M. 2011. Analiza poplav z uporabo satelitskih posnetkov - primer hudourniške poplave v Selški dolini 2007. *Geodetski vestnik* 55, 3: 193–204.
- Liao, T.W. 2008. Enterprise Data Mining: A Review and Research Directions. *Recent Advances In Data Mining Of Enterprise Data: Algorithms and Applications*: 1–109.
- Liew, S.C., Kam, S.P., Tuong, T.P., Chen, P., Minh, V.Q., Lim, H. 1998. Application of multitemporal ERS-1 synthetic aperture radar in delineating rice cropping systems in the Mekong River Delta, Vietnam. *IEEE Transactions on Geoscience and Remote Sensing* 36, 5: 1412–1420.
- Liu, Z., Huang, F., Li, L., Wan, E. 2002. Dynamic monitoring and damage evaluation of flood in north-west Jilin with remote sensing. *International Journal of Remote Sensing* 23, 18: 3669–3679.
- Maimon, O., Rokach, L. 2005. Introduction to supervised method. V: Maimon, O. (ur.), Rokach, L. (ur.). *Data mining and knowledge discovery handbook*. New York, Springer Science & Business: 149–164.
- Marchi, L., Borga, M., Preciso, E., Gaume, E. 2010. Characterisation of selected extreme flash floods in Europe and implications for flood risk management. *Journal of Hydrology* 394: 118–133.

- Martimort, P., Arino, O., Berger, M., Biasutti, R., Carnicero, B., Del Bello, U., Fernandez, V., Gascon, F., Silvestrin, P., Spoto, F., Sy, O. 2007. Sentinel-2 optical high resolution mission for GMES operational services. IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2007: 2677–2680.
- Martimort, P., Fernandez, V., Kirschner, V., Isola, C., Meygret, A. 2012. Sentinel-2 MultiSpectral imager (MSI) and calibration/validation. IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2012: 6999–7002.
- Mason, D.C., Speck, R., Devereux, B., Schumann, G.J.-P., Neal, J.C., Bates, P.D. 2010. Flood Detection in Urban Areas Using TerraSAR-X. IEEE Transactions on Geoscience and Remote Sensing 48, 2: 882–894.
- Mason, D.C., Davenport, I.J., Neal, J.C., Schumann, G. J.-P., Bates, P.D. 2012. Near Real-Time Flood Detection in Urban and Rural Areas Using High-Resolution Synthetic Aperture Radar Images. IEEE Transactions on Geoscience and Remote Sensing 50, 8: 3041–3052.
- Mease, D., Wyner, A.J., Buja, A. 2007. Boosted Classification Trees and Class Probability/Quantile Estimation. Journal of Machine Learning Research 8: 409–439.
- Mikoš, M. 2007. Upravljanje tveganj in nova Evropska direktiva o poplavnih tveganjih. Gradbeni vestnik 56, 11.
- Mikoš, M. 2009. Preventiva nenehno v senci kurative: prof. dr. Matjaž Mikoš, hidrolog. Delo, Gostujoče pero: 51, 300.
- Mitchell, T. 1999. The role of unlabeled data in supervised learning. Proceedings of the 6th International Colloquium on Cognitive Science, San Sebastian, Spain: 1–8.
- Mivšek, E., Žvokelj, B.P., Kete, P., Globokar, T. 2012. Dejanska raba vodnih zemljišč in zemljiški kataster. Geodetski vestnik 56, 4: 663–673.
- MKGP. 2009. Posodobitev grafičnih enot zemljišč kmetijskih gospodarstev (GERK). Novinarska konferenca, 26.11.2009 (dodatno gradivo za medije). Ministrstvo za kmetijstvo, gozdarstvo in prehrano. Služba za odnose z javnostmi in promocijo.
- Mumby, P.J., Green, E.P., Edwards, A.J., Clark C.D. 1999. The cost-effectiveness of remote sensing for tropical coastal resources assessment and management. Journal of Environmental Management 55: 157–166.
- Navulur, K. 2007. Multi-spectral image analysis using the object oriented paradigm. CRC Press, Taylor & Francis Group: 184 str.
- Nigam, K., McCallum, A.K., Thrun, S., Mitchell, T.M. 2000. Text classification from labeled and unlabeled documents using EM. Machine Learning 39, 2/3: 103–134.
- Osman, J., Inglada, J., Christophe, E. 2009. Interactive object segmentation in high resolution satellite images. IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2009: V-48–V-51.
- Oštir, K., Veljanovski, T., Podobnikar, T., Stančič, Z. 2003. Application of satellite remote sensing in natural hazard management: the Mount Mangart landslide case study. International Journal of Remote Sensing 24, 20: 3983–4002.
- Oštir, K. 2006. Daljinsko zaznavanje. Ljubljana, Založba ZRC: 250 str.

- Oštir, K., Marsetič, A., Pehani, P., Perše, M., Veljanovski, T., Zakšek, K., Zaletelj, J., Rodič, T. 2012. Prototype of an automatic near-real-time satellite image processing chain. V: *Small Satellites Systems and Services Symposium, 4S Symposium, Portorož, Slovenija*. ESA: 1–14.
- Pe'eri, S., Philpot, W. 2007. Increasing the Existence of Very Shallow-Water LIDAR Measurements Using the Red-Channel Waveforms. *IEEE Transactions on Geoscience and Remote Sensing* 45, 5: 1217–1223.
- Pehani, P., Kokalj, Ž., Marsetič, A., Oštir, K. 2008. Uporaba satelitskih posnetkov za analizo poplav septembra 2007. V: Perko, D. (ur.), Zorn, M. (ur.), Razpotnik, N. (ur.), Čeh, M. (ur.), Hladnik, D. (ur.), Krevs, M. (ur.), Podobnikar, T. (ur.), Repe, B. (ur.), Šumrada, R. (ur.). *Geografski informacijski sistemi v Sloveniji 2007-2008*. Ljubljana, Geografski inštitut Antona Melika ZRC-SAZU: 117–128.
- Podgorelec, V., Družovec, T.W., Rozman, I. 2005. Some applications of intelligent systems in medicine. *IEEE 3rd International Conference on Computational Cybernetics*: 35–39.
- Podobnikar, T., Mlinar, J. 2006. Izdelava in vzdrževanje digitalnega modela reliefa Slovenije z integracijo obstoječih virov. *Geodetski vestnik* 50, 3: 472–880.
- Podobnikar, T., Szekely, B., Hollaus, M., Roncat, A., Dorninger, P., Briese, C., Melzer, T., Pathe, C., Hofle, B., Pfeifer, N. 2010. Vsestranska uporaba aero-laserskega skeniranja za ugotavljanje nevarnosti zaradi naravnih nesreč na območju Alp. V: Zorn, M. (ur.), Komac, B. (ur.), Pavšek, P. (ur.), Pagon, P. (ur.). *Naravne nesreče 1: Od razumevanja do upravljanja*. Ljubljana, Založba ZRC: 125–137.
- Pohl, C., Van Genderen, J.L. 1998. Multisensor image fusion in remote sensing: concepts, methods and applications. *International Journal of Remote Sensing* 19, 5: 823–854.
- Puech, C., Hostache, R., Raclot, R., Matgen, P. 2009. Quality of flooded plains free water detection using radar images: Towards a methodology for estimation water depth. V: Corbane, C. (ur.), Broglia, M. (ur.), Carrion, D. (ur.), Lemoine, G. (ur.), Pesaresi, M. (ur.). *VALgEO 2009 Proceedings. International workshop on Validation of geo-information products for crisis management, European Communities*: 91–94.
- Pulvirenti, L., Chini, M., Pierdicca, N., Guerriero, L., Ferrazzoli, P. 2011. Flood monitoring using multi-temporal COSMO-SkyMed data: Image segmentation and signature interpretation. *Remote Sensing of Environment* 115, 4: 990–1002.
- Quinlan, J.R. 1986. Introduction of decision trees. *Machine Learning* 1: 81–106.
- Quinlan, J.R. 1992. Learning with continuous classes. V: *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, Hobart, Tasmania. Singapore, World Scientific: 343–348.
- Quinlan, J.R. 1993. *C4.5 Programs for Machine Learning*. San Mateo, CA, Morgan Kaufmann: 270 str.
- Roussopoulos, N., Kelly, S., Vincent, F. 1995. Nearest Neighbor Queries. *Proceedings of the 1995 ACM-SIGMOD International Conference on Management of Data*, San Jose, CA.
- Rusjan, S., Kobold, M., Mikoš, M. 2009. Characteristics of the extreme rainfall event and consequent flash floods in W Slovenia in September 2007. *Natural Hazards and Earth System Sciences* 9, 3: 947–956.

- Sandholt, I., Nyborg, L., Fog, B., Lo, M., Bocoum, O. and Rasmussen, K. 2003. Remote sensing techniques for flood monitoring in the Senegal River Valley. *Danish Journal of Geography* 103: 71–81.
- Scholkopf, B. 1997. *Support Vector Learning*. Munich: Oldenbourg Verlag.
- Schumann, G., Bates, P.D., Horritt, M.S., Matgen, P., Pappenberger, F. 2009. Progress in integration of remote sensing-derived flood extent and stage data and hydraulic models. *Reviews of Geophysics*, 47, 4.
- Segl, K., Kaufmann, H. 2001. Detection of Small Objects From High-Resolution Panchromatic Satellite Imagery Based on Supervised Image Segmentation. *IEEE Transactions on Geoscience and Remote Sensing* 39, 9: 2080–2083.
- Shekhar, S., Vatsavai, R.R., Chawla, S. 2009. Spatial Classification and Prediction Models for Geospatial Data Mining. V: Miller J.H. (ur.), Han, J. (ur.). *Geographic Data Mining and Knowledge Discovery, Second Edition*. CRC Press, Taylor & Francis Group: 117–147.
- Singh, A. 1989. Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing* 10, 6: 989–1003.
- Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., Zehe, E. 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal* 48, 6: 857–880.
- Sivapragasam, C., Liong, S.Y. 2004. Identifying Optimal Training Data Set – A new Approach. *Proceedings of the 6th International Conference on Hydroinformatics*, Singapore: 1621–1628.
- Skidmore, M., Toya, H. 2002. Do natural disasters promote long-run growth? *Economic Inquiry* 40, 4: 664–687.
- Sodnik, J. 2007. *Poplave v septembru 2007 – drobirski tok nad Kropo*. Slovenski vodar 18. Ljubljana.
- Sodnik, J., Mikoš, M. 2011. *Varstvo pred poplavami v Sloveniji*. V: Zorn, M. (ur.), Komac, B. (ur.), Ciglič, R. (ur.), Pavšek, M. (ur.). *Naravne nesreče 2: Neodgovorna odgovornost*. Ljubljana, Založba ZRC: 51–57.
- Stancalie, G., Catana, S., Irimescu, A., Savin, E., Diamandi, A., Hofnar, A., Oancea, S. 2006. Contribution of Earth Observation Data Supplied by the new Satellite Sensors to Flood Management. V: *Transboundary Floods: Reducing Risks Through Flood Management*: 287–304.
http://dx.doi.org/10.1007/1-4020-4902-1_27 (Pridobljeno 26. 3. 2010)
- Štular, B., Kokalj, Ž., Oštir, K., Nuninger, L. 2012. Visualization of lidar-derived relief models for detection of archaeological features. *Journal of Archaeological Science* 39, 11: 3354–3360.
- Tramutoli, V. 2005. Robust Satellite Techniques (RST) for natural and environmental hazards monitoring and mitigation: ten years of successful applications. *The 9th International Symposium on Physical Measurements and Signatures in Remote Sensing*. ISPRS: 792–795.

- Tramutoli, V. 2007. Robust Satellite Techniques (RST) for Natural and Environmental Hazards Monitoring and Mitigation: Theory and Applications, MultiTemp 2007. Fourth International Workshop on the Analysis of Multitemporal Remote Sensing Images.
- UNISDR. 2011. Disasters in numbers, 2011.
http://www.unisdr.org/files/24692_2011disasterstats.pdf (Pridobljeno 28.11.2012.)
- Van Trung, N., Choi, J. H., Won, J. S. 2012. A Land Cover Variation Model of Water Level for the Floodplain of Tonle Sap, Cambodia, Derived From ALOS PALSAR and MODIS Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 99: 1–16.
- Veljanovski, T., Lamovec, P., Pehani, P., Kokalj, Ž., Oštir, K. 2011a. Flooded areas determination using radar satellite images in Slovenia. V: Proc. of the the 7th Conference on Image Information Mining: Geospatial Intelligence from Earth Observation. Ispra, Joint Research Centre, 30 March – 1 April 2011. Ispra, European Union: 141–144.
- Veljanovski, T., Pehani, P., Kokalj, Ž., Oštir, K. 2011b. Zaznavanje poplav s časovno vrsto radarskih satelitskih posnetkov Envisat in Radarsat-2. V: Zorn, M. (ur.), Komac, B. (ur.), Ciglič, R. (ur.), Pašek, M. (ur.). Neodgovorna odgovornost, Naravne nesreče, 2. Ljubljana: Založba ZRC: 81–89.
- Veljanovski, T., Kanjir, U., Oštir, K. 2011c. Objektno usmerjena analiza podatkov daljinskega zaznavanja = Object-based image analysis of remote sensing data. Geod. vestn. 55, 4: 641–688.
- Veljanovski, T., Pehani, P., Lamovec, P., Oštir, K. 2012. Uporabnost podatkov satelitskega in letalskega daljinskega zaznavanja za opazovanje in kartiranje vodnih površin. Geodetski vestnik 56, 4: 786–801.
- Wang, Y., Koopmans, B.N., Pohl, C. 1996. The 1995 Flood in the Netherlands Monitored from Space - a Multisensor Approach. Proceedings of the Second ERS Applications Workshop, London, UK, Dec. 1995.
<http://adsabs.harvard.edu/abs/1996ESASP.383..469W> (Pridobljeno 19. 2. 2010)
- Weirich, F., Blesius, L. 2007. Comparison of satellite and air photo based landslide susceptibility maps. Geomorphology 87, 4: 352–364.
- Weiss, G. M., Provost, F. 2003. Learning when Training Data are costly: The Effect of Class Distribution on Tree Induction. Journal of Artificial Intelligence Research 19: 315–354.
- Westgate, K.N., O'Keefe, P. 1976. Some Definitions of Disaster. Disaster Research Unit Occasional Paper 4. Department of Geography, University of Bradford.
- Witten, I.H., Frank, E. 2005. Data Mining: Practical Machine Learning Tools and Techniques, Second Edition. San Francisco, Morgan Kaufmann: 525 str.
- Yonghua, S., Xiaojuan, L., Huili, G., Wenji, Z., Zhaoning, G. 2007. A study on optical and SAR data fusion for extracting flooded area. International Geoscience and Remote Sensing Symposium IGARSS 2007. IEEE International: 3086–3089.
- Yu, P.S., Chen, S.T., Chang, I.F. 2006. Support vector regression for real-time flood stage forecasting. Journal of Hydrology 328, 3–4: 704–716.

Zha, Y., Gao, J., Ni, S. 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing* 24, 3: 583–594.

Ženko, B. 2003. Izboljšave metode skladanja klasifikatorjev. Magistrsko delo. Univerza v Ljubljani, Fakulteta za računalništvo in informatiko: 60 str.

11 PRILOGE

PRILOGA A.1: ATRIBUTI, UPORABLJENI ZA STROJNO UČENJE NA OBMOČJU ŽELEZNIKOV

Preglednica prikazuje attribute, ki so bili uporabljeni za strojno učenje na območju Železnikov. Prvi stolpec vsebuje oznake atributov, drugi stolpec imena atributov, v tretjem so predstavljeni viri podatkov, iz katerega so atributi pridobljeni, četrti stolpec vsebuje razpon vrednosti posameznega atributa in peti stolpec razlago vrednosti, ki jih ta atribut lahko zavzame.

Oznaka	Atribut	Vir podatka	Razpon vrednosti	Razlaga vrednosti
SPOT B1	Količina odbite zelene svetlobe	SPOT 5 (ms)	65–247	Razpon referenčnih vrednosti: 0–255 (8-bitni zapis digitalnih vrednosti) - 0 pomeni, da se vsa svetloba absorbira, nič svetlobe se ne odbije do sprejemnika na satelitu in - 255 pomeni najvišjo možno vrednost zaznane svetlobe, ki jo senzorji na satelitu lahko zaznajo (večina svetlobe se odbije od zem. površja)
SPOT B1-pansharpen		SPOT 5 (ms), pansharpen	61–254	
SPOT B2	Količina odbite rdeče svetlobe	SPOT 5 (ms)	43–251	
SPOT B2-pansharpen		SPOT 5 (ms), pansharpen	43–254	
SPOT B3	Količina odbite bližnje IR svetlobe	SPOT 5 (ms)	25–161	
SPOT B3-pansharpen		SPOT 5 (ms), pansharpen	30–155	
SPOT B4	Količina odbite kratkovalovne IR svetlobe	SPOT 5 (ms)	24–254	
SPOT B4-pansharpen		SPOT 5 (ms), pansharpen	25–251	
SPOT pan	Količina odbite vidne svetlobe	SPOT 5 (pan)	58–254	
NDVI	Normiran diferencialni vegetacijski indeks	SPOT 5 (ms), pansharpen	–0,7–+0,4	Razpon referenčnih vrednosti od –1 do +1: višje vrednosti pomenijo bolj intenzivno vegetacijo, nižje vrednosti so na območjih brez veg. (npr. voda)
NBI	Novi indeks grajenega okolja	SPOT 5 (ms), pansharpen	0–479	Višje vrednosti so značilne za nerodovitna in pozidana tla, nižje vrednosti za vodna območja
NDBI	Normiran diferencialni indeks grajenega okolja	SPOT 5 (ms), pansharpen	–0,3–+0,4	Razpon referenčnih vrednosti od –1 do +1: pozitivne vrednosti imajo nerodovite in pozidane površine, negativne vrednosti imajo vodna zemljišča
Višina	Nadmorska višina	DMV 0,5 DMV 12,5	436–741 m	Vrednosti predstavljajo nadmorsko višino v metrih
Naklon	Naklon površja	DMV 0,5 DMV 12,5	0,2–49°	Vrednosti predstavljajo spremembo nadmorske višine v smeri največjega upada pobočja (izraženo v stopinjah).
Usmerjenost	Usmerjenost površja	DMV 0,5 DMV 12,5	0–360°	Vrednosti predstavljajo smer največjega upada pobočja. 0° in 360° predstavljata usmerjenost proti severu merjeno v smeri urinega kazalca.
Oddaljenost od vodotokov	Oddaljenost od vodotokov	Mreža vodotokov GKB 50 (linijski objekti sloja vode)	0–394 m	Vrednosti predstavljajo oddaljenost od linijskih vodotokov (reke s pritoki)

PRILOGA A.2: ATRIBUTI, UPORABLJENI ZA STROJNO UČENJE NA OBMOČJU LJUBLJANSKEGA BARJA

Preglednica prikazuje attribute, ki so bili uporabljeni za strojno učenje na območju Ljubljanskega barja. Tako kot v preglednici atributov na območju Železnikov (priloga A.1) prvi stolpec vsebuje oznake atributov, drugi stolpec imena atributov, v tretjem so predstavljeni viri podatkov, iz katerega so atributi pridobljeni, četrti stolpec vsebuje razpon vrednosti posameznega atributa in peti stolpec razlago vrednosti, ki jih ta atribut lahko zavzame.

Oznaka	Atribut	Vir podatka	Razpon vrednosti	Razlaga vrednosti
RapidEye B1	Količina odbite modre svetlobe	RapidEye (ms)	4.360–32.416	Razpon referenčnih vrednosti: 12-bitni zapis digitalnih vrednosti - nizke vrednosti: veliko svetloba se na zem. površju absorbira, nič svetlobe se ne odbije do sprejemnika na satelitu - visoke vrednosti: zem. površje absorbira malo svetlobe, veliko svetlobe se od zem. površja odbije nazaj proti sprejemniku
RapidEye B2	Količina odbite zelene svetlobe	RapidEye (ms)	2.842–31.366	
RapidEye B3	Količina odbite rdeče svetlobe	RapidEye (ms)	1.590–28.205	
RapidEye B4	Količina odboja na zgornjem robu rdeče svetlobe	RapidEye (ms)	1.366–23.755	
RapidEye B5	Količina odbite bližnje IR svetlobe	RapidEye (ms)	878–20.496	
NDVI	Normiran diferencialni vegetacijski indeks	RapidEye (ms)	–0,5–+0,7	Razpon referenčnih vrednosti od –1 do +1: višje vrednosti pomenijo bolj intenzivno vegetacijo, nižje vrednosti so na območjih brez vegetacije (npr. voda)
Višina	Nadmorska višina	DMV 12,5	285–606 m	Vrednosti predstavljajo nadmorsko višino v metrih
Naklon	Naklon površja	DMV 12,5	0–67%	Vrednosti predstavljajo spremembo nadmorske višine v smeri največjega upada pobočja (izraženo v procentih).
Usmerjenost	Usmerjenost površja	DMV 12,5	0–360°	Vrednosti predstavljajo smer največjega upada pobočja. 0° in 360° predstavljata usmerjenost proti severu merjeno v smeri urinega kazalca.
Ukrivljenost	Ukrivljenost površja	DMV 12,5	–6,7– +3,7	Vrednosti predstavljajo stopnjo ukrivljenosti površja. Pozitivne vrednosti označujejo konveksno in negativne vrednosti konkavno ukrivljenost.
Oddaljenost od vodotokov	Oddaljenost od vodotokov	Mreža vodotokov GKB 50 (linijski objekti sloja vode)	2–585 m	Vrednosti predstavljajo oddaljenost od linijskih vodotokov (reke s pritoki in vodnih kanalov)

PRILOGA B: MNOŽICA ODLOČITVENIH DREVES PRIDOBLENIH Z METODO GOZDOV Z 10 DREVESI NA OBMOČJU ŽELEZNIKOV

Priloga B prikazuje izpis odločitvenih dreves, ki so bila pridobljena z metodo gozdov (bagging) v kombinaciji z uporabo algoritma odločitvenih dreves J48 za območje Železnikov. Izpis je prikazan v obliki odločitvenih dreves, kot jih izpiše program Weka.

```
=== Run information ===
Scheme:          weka.classifiers.meta.Bagging -P 100 -S 1 -I 10 -W
weka.classifiers.trees.J48 -- -C 0.25 -M 2
Relation:        atributi4_145tock-weka.filters.unsupervised.attribute.Remove-
R1-weka.filters.unsupervised.attribute.Remove-R2-3-
weka.filters.unsupervised.attribute.Remove-R4-
weka.filters.unsupervised.attribute.Remove-R6
Instances:       145
Attributes:      17
                  poplave
                  visina
                  naklon
                  usmerjenost
                  oddaljenost
                  NDVI_pansharpen
                  B1
                  B2
                  B3
                  B4
                  B1pansharp
                  B2pansharp
                  B3pansharp
                  B4pansharp
                  SPOT_pan
                  NBI
                  NDBI
Test mode:       10-fold cross-validation

=== Classifier model (full training set) ===
All the base classifiers:

J48 pruned tree
-----
NDVI_pansharpen <= -0.283951
|   B4 <= 112
|   |   B4pansharp <= 105: 1 (40.0)
|   |   B4pansharp > 105
|   |   |   B1pansharp <= 167: 1 (3.0)
|   |   |   B1pansharp > 167: 0 (3.0)
|   B4 > 112: 0 (6.0/1.0)
NDVI_pansharpen > -0.283951
|   B2pansharp <= 121: 0 (69.0)
|   B2pansharp > 121
|   |   SPOT_pan <= 132: 1 (9.0/1.0)
|   |   SPOT_pan > 132: 0 (15.0)

Number of Leaves :      7
```

Size of the tree : 13

J48 pruned tree

```
NDVI_pansharpen <= -0.142857
| SPOT_pan <= 160
| | oddaljenost <= 39: 1 (39.0/1.0)
| | oddaljenost > 39
| | | naklon <= 1.538526: 1 (9.0)
| | | naklon > 1.538526: 0 (9.0)
| SPOT_pan > 160: 0 (17.0)
NDVI_pansharpen > -0.142857: 0 (71.0)
```

Number of Leaves : 5

Size of the tree : 9

J48 pruned tree

```
B1pansharp <= 131: 0 (67.0)
B1pansharp > 131
| B4pansharp <= 120
| | SPOT_pan <= 135: 1 (36.0)
| | SPOT_pan > 135
| | | B3pansharp <= 93
| | | | usmerjenost <= 310: 1 (19.0/1.0)
| | | | usmerjenost > 310: 0 (2.0)
| | | B3pansharp > 93: 0 (3.0)
| B4pansharp > 120: 0 (18.0)
```

Number of Leaves : 6

Size of the tree : 11

J48 pruned tree

```
B1pansharp <= 131: 0 (66.0)
B1pansharp > 131
| B3pansharp <= 102
| | oddaljenost <= 74.788033: 1 (46.0)
| | oddaljenost > 74.788033
| | | NDBI <= 0.148148: 1 (9.0/1.0)
| | | NDBI > 0.148148: 0 (3.0)
| B3pansharp > 102: 0 (21.0)
```

Number of Leaves : 5

Size of the tree : 9

J48 pruned tree

```
NDVI_pansharpen <= -0.171548
| B4pansharp <= 116
| | B3 <= 38: 0 (6.0)
| | B3 > 38
| | | oddaljenost <= 39: 1 (34.0)
| | | oddaljenost > 39
| | | | naklon <= 1.538526: 1 (12.0/1.0)
| | | | naklon > 1.538526: 0 (10.0/1.0)
| B4pansharp > 116: 0 (9.0)
NDVI_pansharpen > -0.171548: 0 (74.0)
```

Number of Leaves : 6
Size of the tree : 11

J48 pruned tree

```
-----  
NBI <= 96: 0 (63.0)  
NBI > 96  
| B4pansharp <= 116  
| | oddaljenost <= 39.5: 1 (45.0/1.0)  
| | | oddaljenost > 39.5  
| | | SPOT_pan <= 126: 0 (5.0)  
| | | SPOT_pan > 126  
| | | | usmerjenost <= 305: 1 (11.0/1.0)  
| | | | usmerjenost > 305: 0 (3.0)  
| B4pansharp > 116: 0 (18.0)
```

Number of Leaves : 6
Size of the tree : 11

J48 pruned tree

```
-----  
NDVI_pansharpen <= -0.142857  
| B4pansharp <= 122  
| | B4pansharp <= 45: 0 (3.0)  
| | B4pansharp > 45  
| | | NDBI <= 0.161765  
| | | | B3 <= 90: 1 (47.0)  
| | | | B3 > 90: 0 (3.0/1.0)  
| | | NDBI > 0.161765  
| | | | naklon <= 4.883837: 0 (4.0)  
| | | | naklon > 4.883837: 1 (2.0)  
| B4pansharp > 122: 0 (14.0)  
NDVI_pansharpen > -0.142857: 0 (72.0)
```

Number of Leaves : 7
Size of the tree : 13

J48 pruned tree

```
-----  
NDVI_pansharpen <= -0.196653  
| B4pansharp <= 120  
| | B3 <= 38: 0 (5.0)  
| | B3 > 38  
| | | oddaljenost <= 40.140377: 1 (41.0)  
| | | oddaljenost > 40.140377  
| | | | B4pansharp <= 102: 1 (11.0/1.0)  
| | | | B4pansharp > 102: 0 (5.0)  
| B4pansharp > 120: 0 (11.0)  
NDVI_pansharpen > -0.196653: 0 (72.0/1.0)
```

Number of Leaves : 6
Size of the tree : 11

J48 pruned tree

```
NBI <= 96: 0 (60.0)
NBI > 96
|  B4pansharp <= 120
|  |  oddaljenost <= 39: 1 (43.0)
|  |  oddaljenost > 39
|  |  |  naklon <= 2.489273
|  |  |  |  usmerjenost <= 237: 1 (16.0)
|  |  |  |  usmerjenost > 237: 0 (4.0/1.0)
|  |  |  naklon > 2.489273: 0 (6.0)
|  B4pansharp > 120: 0 (16.0)
```

```
Number of Leaves   :    6
Size of the tree   :   11
```

J48 pruned tree

```
NDVI_pansharpen <= -0.242291
|  B4 <= 111
|  |  visina <= 467.730011
|  |  |  oddaljenost <= 74.788033: 1 (45.0)
|  |  |  oddaljenost > 74.788033
|  |  |  |  visina <= 448.290009: 1 (3.0)
|  |  |  |  visina > 448.290009: 0 (2.0)
|  |  |  visina > 467.730011: 0 (6.0/1.0)
|  B4 > 111: 0 (10.0)
NDVI_pansharpen > -0.242291: 0 (79.0/1.0)
```

```
Number of Leaves   :    6
Size of the tree   :   11
```

PRILOGA C: MNOŽICA ODLOČITVENIH DREVES PRIDOBLENIH Z METODO NAKLJUČNIH GOZDOV Z 10 DREVESI NA OBMOČJU ŽELEZNIKOV

Priloga C prikazuje izpis odločitvenih dreves, ki so bila pridobljena z metodo naključnih gozdov (random forests) za območje Železnikov. Izpis odločitvenih dreves je prikazan v programskem jeziku Python, kot nam ga poda program Clus.

```
# Python code of the trees in the ensemble

#Model 1
def clus_tree_1( visina, naklon, usmerjenost, oddaljenost, NDVI-pansharpen,
B1, B2, B3, B4, B1pansharp, B2pansharp, B3pansharp, B4pansharp, SPOT pan,
NBI, NDBI ):
    if B1 > 112.0:
        if B3pansharp > 102.0:
            return [0]
        else:
            if oddaljenost > 76.922035:
                if NBI > 155.149246:
                    return [0]
                else:
                    return [1]
            else:
                return [1]
    else:
        if B1pansharp > 126.0:
            if NDVI-pansharpen > -0.242291:
                return [0]
            else:
                return [1]
        else:
            return [0]

#Model 2
def clus_tree_2( visina, naklon, usmerjenost, oddaljenost, NDVI-pansharpen,
B1, B2, B3, B4, B1pansharp, B2pansharp, B3pansharp, B4pansharp, SPOT pan,
NBI, NDBI ):
    if NDVI-pansharpen > -0.171548:
        return [0]
    else:
        if oddaljenost > 37.792194:
            if NDBI > 0.097222:
                return [0]
            else:
                if B3pansharp > 68.0:
                    if B3 > 86.0:
                        return [0]
                    else:
                        return [1]
                else:
                    return [0]
        else:
            if B3pansharp > 90.0:
                return [0]
            else:
                return [1]
```

```
#Model 3
def clus_tree_3( visina, naklon, usmerjenost, oddaljenost, NDVI-pansharpen,
B1, B2, B3, B4, B1pansharp, B2pansharp, B3pansharp, B4pansharp, SPOT pan,
NBI, NDBI ):
    if visina > 467.850006:
        return [0]
    else:
        if B4pansharp > 114.0:
            return [0]
        else:
            if B1pansharp > 131.0:
                if NBI > 179.520004:
                    if B3 > 66.0:
                        return [1]
                    else:
                        if SPOT pan > 133.0:
                            return [0]
                        else:
                            return [1]
            else:
                return [1]
        else:
            return [0]

#Model 4
def clus_tree_4( visina, naklon, usmerjenost, oddaljenost, NDVI-pansharpen,
B1, B2, B3, B4, B1pansharp, B2pansharp, B3pansharp, B4pansharp, SPOT pan,
NBI, NDBI ):
    if NDVI-pansharpen > -0.142857:
        return [0]
    else:
        if SPOT pan > 159.0:
            if NDVI-pansharpen > -0.361538:
                return [0]
            else:
                if visina > 450.76001:
                    return [1]
                else:
                    return [0]
        else:
            if B1pansharp > 119.0:
                if B2 > 104.0:
                    return [1]
                else:
                    if NDVI-pansharpen > -0.365079:
                        return [0]
                    else:
                        return [1]
            else:
                return [0]

#Model 5
def clus_tree_5( visina, naklon, usmerjenost, oddaljenost, NDVI-pansharpen,
B1, B2, B3, B4, B1pansharp, B2pansharp, B3pansharp, B4pansharp, SPOT pan,
NBI, NDBI ):
    if B1pansharp > 126.0:
        if B4pansharp > 111.0:
            return [0]
        else:
            if B3 > 76.0:
                if NDVI-pansharpen > -0.198157:
```

```
        return [0]
    else:
        return [1]
else:
    if B1pansharp > 169.0:
        if B1 > 131.0:
            return [1]
        else:
            return [0]
    else:
        return [1]
else:
    return [0]

#Model 6
def clus_tree_6( visina, naklon, usmerjenost, oddaljenost, NDVI-pansharpen,
B1, B2, B3, B4, B1pansharp, B2pansharp, B3pansharp, B4pansharp, SPOT pan,
NBI, NDBI ):
    if NDVI-pansharpen > -0.319149:
        if NBI > 101.281044:
            if B1pansharp > 163.0:
                return [0]
            else:
                if NDVI-pansharpen > -0.142857:
                    return [0]
                else:
                    return [1]
        else:
            return [0]
    else:
        if B1 > 174.0:
            return [0]
        else:
            if NDBI > 0.148148:
                if B4 > 68.0:
                    if visina > 446.549988:
                        return [0]
                    else:
                        return [1]
                else:
                    return [1]
            else:
                return [1]
        else:
            return [1]

#Model 7
def clus_tree_7( visina, naklon, usmerjenost, oddaljenost, NDVI-pansharpen,
B1, B2, B3, B4, B1pansharp, B2pansharp, B3pansharp, B4pansharp, SPOT pan,
NBI, NDBI ):
    if B2pansharp > 140.0:
        if B4pansharp > 106.0:
            if B1pansharp > 148.0:
                if B2pansharp > 164.0:
                    return [0]
                else:
                    return [1]
            else:
                return [1]
        else:
            return [1]
    else:
        if NBI > 223.670883:
            return [0]
        else:
```

```

        return [1]
    else:
        if NDVI-pansharpen > -0.365079:
            return [0]
        else:
            return [1]

#Model 8
def clus_tree_8( visina, naklon, usmerjenost, oddaljenost, NDVI-pansharpen,
B1, B2, B3, B4, B1pansharp, B2pansharp, B3pansharp, B4pansharp, SPOT pan,
NBI, NDBI ):
    if NBI > 81.310921:
        if B4pansharp > 120.0:
            return [0]
        else:
            if B3pansharp > 90.0:
                if visina > 471.51001:
                    return [1]
                else:
                    return [0]
            else:
                if visina > 467.730011:
                    return [0]
                else:
                    return [1]
    else:
        return [0]

#Model 9
def clus_tree_9( visina, naklon, usmerjenost, oddaljenost, NDVI-pansharpen,
B1, B2, B3, B4, B1pansharp, B2pansharp, B3pansharp, B4pansharp, SPOT pan,
NBI, NDBI ):
    if NDVI-pansharpen > -0.242291:
        if B1pansharp > 131.0:
            if oddaljenost > 22.102036:
                return [0]
            else:
                return [1]
        else:
            return [0]
    else:
        if B4pansharp > 105.0:
            if visina > 446.01001:
                return [0]
            else:
                if B4pansharp > 114.0:
                    return [0]
                else:
                    return [1]
        else:
            if B1 > 164.0:
                return [0]
            else:
                if B1 > 79.0:
                    return [1]
                else:
                    return [0]

```

```
#Model 10
def clus_tree_10( visina, naklon, usmerjenost, oddaljenost, NDVI-
pansharpen, B1, B2, B3, B4, B1pansharp, B2pansharp, B3pansharp, B4pansharp,
SPOT pan, NBI, NDBI ):
    if NBI > 101.281044:
        if B3pansharp > 102.0:
            return [0]
        else:
            if NDBI > 0.21393:
                return [0]
            else:
                if SPOT pan > 168.0:
                    return [0]
                else:
                    if NDVI-pansharpen > -0.171548:
                        return [1]
                    else:
                        if B3pansharp > 71.0:
                            return [1]
                        else:
                            if B4pansharp > 97.0:
                                return [0]
                            else:
                                return [1]
    else:
        return [0]
```

PRILOGA D: MNOŽICA ODLOČITVENIH DREVES PRIDOBLENIH Z METODO GOZDOV Z 10 DREVESI NA OBMOČJU LJUBLJANSKEGA BARJA

Priloga D prikazuje izpis odločitvenih dreves, ki so bila pridobljena z metodo gozdov (bagging) v kombinaciji z uporabo algoritma odločitvenih dreves J48 za območje Ljubljanskega barja. Izpis je prikazan v obliki odločitvenih dreves, kot jih izpiše program Weka.

```
=== Run information ===
Scheme:          weka.classifiers.meta.Bagging  -P 100  -S 1  -I 10  -W
weka.classifiers.trees.J48 -- -C 0.25 -M 2
Relation:        Tocke_ucenje-weka.filters.unsupervised.attribute.Remove-R1,3-6
Instances:       201
Attributes:      13
                  voda
                  DMV
                  naklon
                  usmerjenost
                  ukrivljenost
                  B1-blue
                  B2-green
                  B3-red
                  B4-red_edge
                  B5-near_IR
                  NDVI
                  SkyView
                  oddaljenost_vod
Test mode:       10-fold cross-validation
```

```
=== Classifier model (full training set) ===
All the base classifiers:
```

J48 pruned tree

```
-----
DMV <= 289.338348
|  NDVI <= 0.363431: 1 (37.0/2.0)
|  NDVI > 0.363431: 0 (8.0)
DMV > 289.338348: 0 (156.0/1.0)
```

```
Number of Leaves  :    3
Size of the tree  :    5
```

J48 pruned tree

```
-----
DMV <= 289.338348
|  B5-near_IR <= 4050: 1 (35.0)
|  B5-near_IR > 4050
|  |  usmerjenost <= 334: 0 (15.0/1.0)
|  |  usmerjenost > 334: 1 (3.0)
DMV > 289.338348: 0 (148.0)
```

```
Number of Leaves  :    4
Size of the tree  :    7
```

J48 pruned tree

```
B5-near_IR <= 1972: 1 (29.0/1.0)
B5-near_IR > 1972
|   DMV <= 289.338348
|   |   B5-near_IR <= 6141
|   |   |   ukrivljenost <= 0.023984: 1 (6.0)
|   |   |   ukrivljenost > 0.023984: 0 (2.0)
|   |   B5-near_IR > 6141: 0 (12.0)
|   DMV > 289.338348: 0 (152.0)
```

Number of Leaves : 5

Size of the tree : 9

J48 pruned tree

```
B5-near_IR <= 1914: 1 (26.0/1.0)
B5-near_IR > 1914
|   DMV <= 289.785461
|   |   B1-blue <= 5051: 0 (15.0)
|   |   B1-blue > 5051
|   |   |   B5-near_IR <= 6123: 1 (11.0)
|   |   |   B5-near_IR > 6123: 0 (4.0)
|   DMV > 289.785461: 0 (145.0)
```

Number of Leaves : 5

Size of the tree : 9

J48 pruned tree

```
B5-near_IR <= 1972: 1 (25.0/1.0)
B5-near_IR > 1972
|   DMV <= 289.338348
|   |   B1-blue <= 5066: 0 (10.0)
|   |   B1-blue > 5066
|   |   |   B1-blue <= 10301: 1 (8.0)
|   |   |   B1-blue > 10301: 0 (3.0)
|   DMV > 289.338348: 0 (155.0)
```

Number of Leaves : 5

Size of the tree : 9

J48 pruned tree

```
DMV <= 289.785461
|   B5-near_IR <= 3495: 1 (27.0)
|   B5-near_IR > 3495
|   |   B3-red <= 2741: 0 (16.0)
|   |   B3-red > 2741
|   |   |   DMV <= 288.372742: 0 (3.0)
|   |   |   DMV > 288.372742: 1 (3.0)
|   DMV > 289.785461: 0 (152.0)
```

Number of Leaves : 5

Size of the tree : 9

J48 pruned tree

```
DMV <= 289.338348
|   B5-near_IR <= 6288: 1 (35.0/2.0)
|   B5-near_IR > 6288: 0 (8.0)
DMV > 289.338348: 0 (158.0)
```

Number of Leaves : 3

Size of the tree : 5

J48 pruned tree

```
B5-near_IR <= 1972: 1 (20.0)
B5-near_IR > 1972
|   DMV <= 289.785461
|   |   B1-blue <= 5059: 0 (17.0)
|   |   B1-blue > 5059
|   |   |   B5-near_IR <= 6141: 1 (10.0)
|   |   |   B5-near_IR > 6141: 0 (5.0)
|   |   DMV > 289.785461: 0 (149.0)
```

Number of Leaves : 5

Size of the tree : 9

J48 pruned tree

```
DMV <= 288.811523
|   B5-near_IR <= 6627: 1 (27.0)
|   B5-near_IR > 6627: 0 (3.0)
DMV > 288.811523
|   DMV <= 289.785461
|   |   DMV <= 289.774567: 0 (18.0/1.0)
|   |   DMV > 289.774567: 1 (2.0)
|   DMV > 289.785461: 0 (151.0)
```

Number of Leaves : 5

Size of the tree : 9

J48 pruned tree

```
DMV <= 289.338348
|   B5-near_IR <= 4139: 1 (36.0)
|   B5-near_IR > 4139
|   |   usmerjenost <= 334: 0 (12.0/1.0)
|   |   usmerjenost > 334: 1 (2.0)
DMV > 289.338348: 0 (151.0)
```

Number of Leaves : 4

Size of the tree : 7

PRILOGA E: MNOŽICA ODLOČITVENIH DREVES PRIDOBLENIH Z METODO NAKLJUČNIH GOZDOV Z 10 DREVESI NA OBMOČJU LJUBLJANSKEGA BARJA

Priloga E prikazuje izpis odločitvenih dreves, ki so bila pridobljena z metodo naključnih gozdov (random forests) za območje Ljubljanskega barja. Izpis odločitvenih dreves je prikazan v programskem jeziku Python, kot nam ga poda program Clus.

```
# Python code of the trees in the ensemble

#Model 1
def clus_tree_1( visina, naklon, usmerjenost, ukrivljenost, B1, B2, B3, B4,
B5, ndvi_ms, oddaljenost ):
    if ndvi_ms > -0.191099:
        if B5 > 1914.0:
            if usmerjenost > 4.0:
                if B5 > 2160.0:
                    if usmerjenost > 346.0:
                        return [0]
                    else:
                        if visina > 288.139893:
                            return [0]
                        else:
                            if visina > 287.746491:
                                return [1]
                            else:
                                return [0]
                        else:
                            return [0]
                    else:
                        return [1]
                else:
                    return [1]
            else:
                return [1]
        else:
            return [1]

#Model 2
def clus_tree_2( visina, naklon, usmerjenost, ukrivljenost, B1, B2, B3, B4,
B5, ndvi_ms, oddaljenost ):
    if visina > 288.139893:
        if naklon > 0.683117:
            return [0]
        else:
            if usmerjenost > 332.0:
                return [1]
            else:
                if B5 > 3263.0:
                    return [0]
                else:
                    return [1]
    else:
        if B4 > 2960.0:
            if B1 > 5087.0:
                return [0]
            else:
                return [1]
        else:
            return [1]
```

```
#Model 3
def clus_tree_3( visina, naklon, usmerjenost, ukrivljenost, B1, B2, B3, B4,
B5, ndvi_ms, oddaljenost ):
    if B4 > 2293.0:
        if naklon > 0.305611:
            if B4 > 3154.0:
                return [0]
            else:
                if B4 > 3103.0:
                    return [1]
                else:
                    return [0]
        else:
            if visina > 289.338348:
                return [0]
            else:
                return [1]
    else:
        if ukrivljenost > -0.194727:
            if ukrivljenost > 0.191113:
                if usmerjenost > 24.0:
                    return [1]
                else:
                    return [0]
            else:
                return [1]
        else:
            return [0]

#Model 4
def clus_tree_4( visina, naklon, usmerjenost, ukrivljenost, B1, B2, B3, B4,
B5, ndvi_ms, oddaljenost ):
    if B5 > 1914.0:
        if usmerjenost > 341.0:
            if usmerjenost > 354.0:
                return [0]
            else:
                return [1]
        else:
            if B4 > 2293.0:
                return [0]
            else:
                if naklon > 0.683117:
                    return [0]
                else:
                    return [1]
    else:
        return [1]

#Model 5
def clus_tree_5( visina, naklon, usmerjenost, ukrivljenost, B1, B2, B3, B4,
B5, ndvi_ms, oddaljenost ):
    if B5 > 1972.0:
        if naklon > 0.187167:
            if visina > 289.338348:
                return [0]
            else:
                if B3 > 2465.0:
                    return [1]
                else:
```

```

        return [0]
    else:
        if naklon > 0.184671:
            return [1]
        else:
            return [0]
    else:
        if visina > 288.450501:
            return [0]
        else:
            return [1]

#Model 6
def clus_tree_6( visina, naklon, usmerjenost, ukrivljenost, B1, B2, B3, B4,
B5, ndvi_ms, oddaljenost ):
    if visina > 289.026337:
        if B4 > 2122.0:
            return [0]
        else:
            if visina > 289.785461:
                return [0]
            else:
                return [1]
    else:
        if oddaljenost > 17.088007:
            if ndvi_ms > 0.165707:
                if naklon > 0.17817:
                    return [0]
                else:
                    return [1]
            else:
                return [1]
        else:
            return [0]

#Model 7
def clus_tree_7( visina, naklon, usmerjenost, ukrivljenost, B1, B2, B3, B4,
B5, ndvi_ms, oddaljenost ):
    if ndvi_ms > -0.216168:
        if B4 > 2293.0:
            if B5 > 5772.0:
                return [0]
            else:
                if usmerjenost > 293.0:
                    if B1 > 6189.0:
                        return [1]
                    else:
                        return [0]
                else:
                    return [0]
        else:
            if ukrivljenost > -0.29877:
                if naklon > 1.430499:
                    return [0]
                else:
                    return [1]
            else:
                return [0]
    else:
        return [1]
```

```
#Model 8
def clus_tree_8( visina, naklon, usmerjenost, ukrivljenost, B1, B2, B3, B4,
B5, ndvi_ms, oddaljenost ):
    if visina > 289.026337:
        return [0]
    else:
        if B4 > 2293.0:
            if B1 > 5066.0:
                if B2 > 6049.0:
                    return [0]
                else:
                    return [1]
            else:
                return [0]
        else:
            return [1]

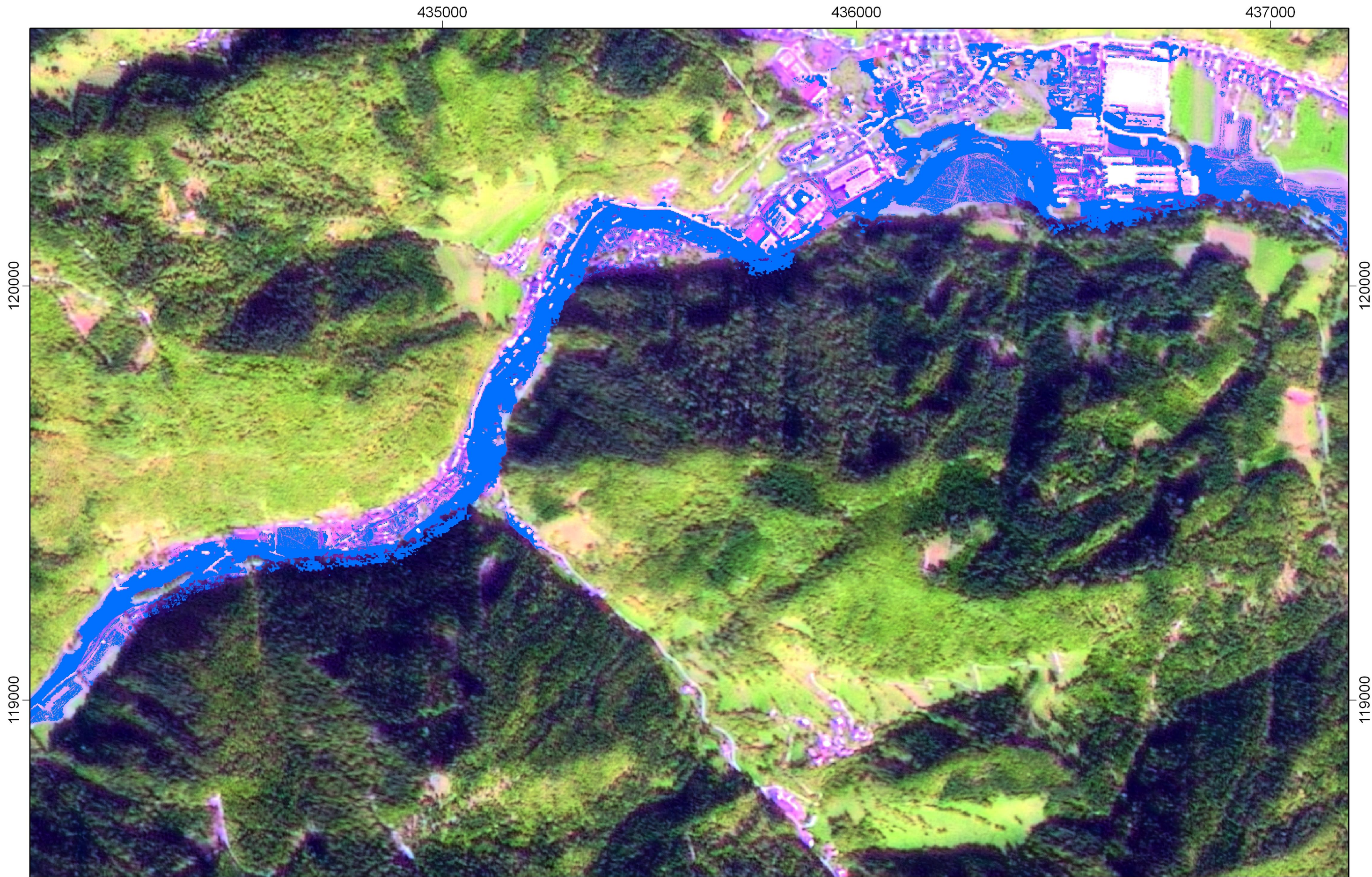
#Model 9
def clus_tree_9( visina, naklon, usmerjenost, ukrivljenost, B1, B2, B3, B4,
B5, ndvi_ms, oddaljenost ):
    if B5 > 1475.0:
        if visina > 289.785461:
            return [0]
        else:
            if B3 > 2241.0:
                if naklon > 0.683117:
                    return [0]
                else:
                    if B5 > 5772.0:
                        return [0]
                    else:
                        return [1]
            else:
                return [0]
    else:
        return [1]

#Model 10
def clus_tree_10( visina, naklon, usmerjenost, ukrivljenost, B1, B2, B3,
B4, B5, ndvi_ms, oddaljenost ):
    if B5 > 1972.0:
        return [0]
    else:
        if visina > 288.450501:
            return [0]
        else:
            return [1]
```

PRILOGA F: KARTI POPLAVLJENIH POVRŠIN NA OBMOČJU ŽELEZNIKOV IN NA OBMOČJU LJUBLJANSKEGA BARJA

Priloga F.1: Karta poplavljenih površin na območju Železnikov

Priloga F.2: Karta poplavljenih površin na območju Ljubljanskega barja

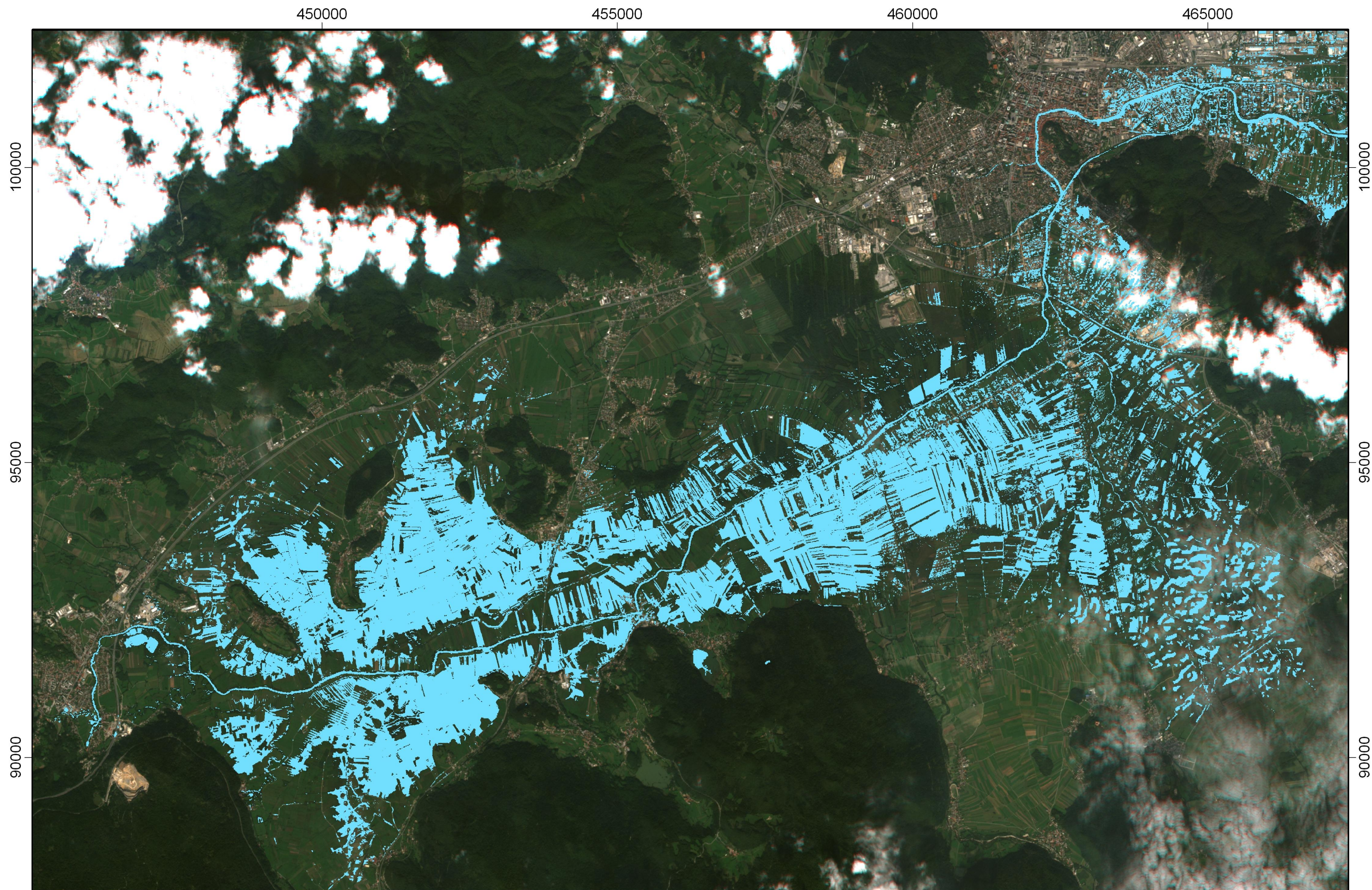


0 0.25 0.5 km

Karta poplavljenih površin na območju Železnikov

Učenje - algoritem: J48, učni vzorec: 145 točk

Natančnost klasifikacije poplavljenih površin - izdelovalec: 98 %, uporabnik: 91 %



0 2.5 5 km

Karta poplavljenih površin na Ljubjanskem barju

Učenje - algoritem: J48, učni vzorec: 201 točk

Natančnost klasifikacije poplavljenih površin - izdelovalec: 84 %, uporabnik: 93 %